

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2004, 14:37:17 ; Search time 39 Seconds
(without alignments)
2281.439 Million cell updates/sec

Title: PCT-US03-12556-2
Perfect score: 1522
Sequence: 1 MKKXVLAAAMVALAPVQS.....BAQRTAGYIHLWFTYYNR 282

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_25:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phase:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1257.5	82.6	283	052864	O52864 bacillus th
2	1257.5	82.6	283	Q81HW1	Q81HW1 bacillus ce
3	1251.5	82.2	283	Q81V27	Q81V27 bacillus an
4	1250.5	82.2	283	Q84FI6	Q84FI6 bacillus ce
5	482.5	31.7	275	2	Q84DK1 listeria se
6	448	29.4	242	2	Q83TL2 listeria mo
7	442	29.0	242	2	Q84DK4 listeria mo
8	442	29.0	242	2	Q83U53 listeria mo
9	440	28.9	242	2	Q84DK0 listeria mo
10	437	28.7	242	2	Q84DK5 listeria mo
11	437	28.7	242	2	Q84DK3 listeria mo
12	411	27.0	236	2	Q84DK6 listeria mo
13	411	27.0	242	2	Q84DK2 listeria mo
14	357.5	23.5	223	2	Q84DK7 listeria mo
15	325	21.4	154	2	Q8VMY6 listeria iv
16	314.5	20.7	178	2	Q48580 listeria iv

17	230	15.1	398	2	Q8GCV3	Q8GCV3 clostridium
18	219	14.4	370	2	Q83Z25	Q83Z25 clostridium
19	166	10.9	272	16	Q97TF6	Q97TF6 clostridium
20	146	9.6	292	17	Q8PY92	Q8PY92 methanosarc
21	123	8.1	393	17	Q8TZ39	Q8TZ39 methanopyru
22	117.5	7.7	242	16	Q896L3	Q896L3 clostridium
23	111	7.3	1042	13	Q9DEG5	Q9DEG5 gallus gall
24	111	7.3	2402	13	Q9DEH4	Q9DEH4 gallus gall
25	109.5	7.2	245	16	Q97KA0	Q97KA0 clostridium
26	108.5	7.1	2140	16	Q97RY6	Q97RY6 streptococc
27	107.5	7.1	325	17	Q8TWV6	Q8TWV6 methanopyru
28	107	7.0	2144	2	Q8S4M8	Q8S4M8 streptococc
29	107	7.0	2144	16	Q8DQF7	Q8DQF7 streptococc
30	106.5	7.0	904	13	Q9DEG4	Q9DEG4 gallus gall
31	106	7.0	607	5	Q8ICW0	Q8ICW0 plasmodium
32	105	6.9	4144	6	Q8WN22	Q8WN22 canis famil
33	103	6.8	625	16	Q97J52	Q97J52 clostridium
34	102.5	6.7	470	16	Q8YX86	Q8YX86 anabaena sp
35	102.5	6.7	956	5	Q97238	Q97238 plasmodium
36	102	6.7	1818	16	Q8EVF1	Q8EVF1 mycoplasma
37	101.5	6.6	570	10	Q24041	Q24041 liriodendro
38	100.5	6.6	477	5	Q25267	Q25267 leishmania
39	100.5	6.6	1740	5	Q81JK9	Q81JK9 plasmodium
40	100	6.6	957	16	Q8EWU3	Q8EWU3 mycoplasma
41	99	6.5	273	16	Q9PHU3	Q9PHU3 xylella fas
42	98.5	6.5	378	5	Q9GNZ4	Q9GNZ4 leishmania
43	98	6.4	352	16	Q88AF4	Q88AF4 pseudomonas
44	98	6.4	472	16	Q8CVP1	Q8CVP1 streptococc
45	98	6.4	988	15	Q8A3K2	Q8A3K2 bacteroides

ALIGNMENTS

RESULT 1

O52864 PRELIMINARY; PRT; 283 AA.
AC O52864
DT 01-JUN-1998 (TREMREL. 06, Created)
DT 01-JUN-1998 (TREMREL. 06, Last sequence update)
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
DE Phosphatidyl-degrading phospholipase C.
GN PLCA.
OS Bacillus thuringiensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1428;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=Bt13;
RX PubMed=9732531;
RA Loevgren A., Carlson C.R., Eskils K., Kolsto A.B.;
RT "Localization of putative virulence genes on a physical map of the
Bacillus thuringiensis subsp. Gelechia chromosome.";
RL Curr. Microbiol. 37:245-250(1998).
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=Bt13;
RA Loevgren A.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y16268; CAA76148.1; -;
DR HSSP; P09598; 1AH7.
DR GO; GO:0004629; F:phospholipase C activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR008947; PLC_Nuclease.
DR InterPro; IPR001531; Zn_dep_PLPC.
DR Pfam; PF00882; Zn_dep_PLPC; 1.
DR PRINTS; PR00479; PREPHLIPASEC.
DR ProDom; PD003946; Zn_dep_PLPC; 1.
DR PROSITE; PS00384; PROKAR_ZN_DEPEND_PLPC; 1.
SQ SEQUENCE 283 AA; 32387 MW; 4D72E6723F8A68FF CRC64;

Query Match 82.6%; Score 1257.5; DB 2; Length 283;
Best Local Similarity 80.2%; Pred. NO. 3.6e-90;

Qy 240 AVSQYADKRAEAVTPVTGKRLMEQRTVAGYVHWFDTYVNR 282
|||||
Db 241 AVSQYADKRAEAVTPVTGKRLMDAQRVTAGYVHWFDTYVNR 283
|||||

RESULT 4

Q84F16 PRELIMINARY; PRT; 283 AA.
ID Q84F16;
AC Q84F16;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Phosphatidylcholine-specific phospholipase C.
GN PC-PLC.
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1396;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=569;
RA Pomerantsev A.P., Kalnin K.V., Leppla S.H.;
RT "Bacillus cereus pc-plc and sph genes.";
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY195600; AAO0751.1; --
DR GO; GO:0004629; F:phospholipase C activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR008947; PLC Nuclease.
DR Pfam; PF00882; Zn dep PLPC; 1.
DR PRINTS; PR00479; PRPHLPASEC.
DR PRODOM; PD003946; Zn dep PLPC; 1.
DR PROSITE; PS00384; PROKAR_ZN_DEPEND_PLPC; 1.
SQ SEQUENCE 283 AA; 32450 MW; 289787200D5B9FC8 CRC64;

Query Match 82.2%; Score 1250.5; DB 2; Length 283;
Best Local Similarity 79.9%; Pred. No. 1.3e-89;
Matches 226; Conservative 24; Mismatches 32; Indels 1; Gaps 1;

Qy 1 MKKKVLALAAAVQSVVFAQTNNSESPAPIL-RWSAEDKHNEGINSHLWTVNRAI 59
|||||
Db 1 MKKKVLALAAAVQSVVFAQTNNSESPAPIL-RWSAEDKHNEGINSHLWTVNRAI 60
|||||
Qy 60 DIMSRNTTIVNPNETALINEWRADLENGIYSADYENPYDNTSYASHFYDPTGTITYPF 119
|||||
Db 61 DIMSRNTLVQDRVALLNEWRTELENGIYAADYENPYDNTSYASHFYDPTGTITYPF 120
|||||
Qy 120 AKHAKETGAKYFNLAGQAYQNDQQOAPFYGLSLHYLGDVNPMPHAANFTNLSYPMGFH 179
|||||
Db 121 AKHAKETGAKYFNLAGQAYQNDQQOAPFYGLSLHYLGDVNPMPHAANFTNLSYPMGFH 180
|||||
Qy 180 SKYENFVDTIKNNYIVSDNSGYNWKGANPEDWIEGAAVAAKQDYPGVVNDTTKDFVKA 239
|||||
Db 181 SKYENFVDTIKNNYIVSDNSGYNWKGANPEDWIEGAAVAAKQDYPGVVNDTTKDFVKA 240
|||||
Qy 240 AVSQYADKRAEAVTPVTGKRLMEQRTVAGYVHWFDTYVNR 282
|||||
Db 241 AVSQYADKRAEAVTPVTGKRLMDAQRVTAGYVHWFDTYVNR 283
|||||

RESULT 5

Q84DK1 PRELIMINARY; PRT; 275 AA.
ID Q84DK1;
AC Q84DK1;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Phospholipase (Fragment).
GN PLCB.
OS Listeria seeligeri.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1640;
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE=22535694; PubMed=12648840;
RA Schmid M., Walcher M., Bubert A., Wagner M., Wagner M.,
Schleifer K.-H.;
RT "Nucleic acid-based, cultivation-independent detection of Listeria
app. and genotypes of L. monocytogenes.";
RL FEMS Immunol. Med. Microbiol. 35:215-225(2003).
DR EMBL; AY150839; AAO19486.1; --
DR GO; GO:0004629; F:phospholipase C activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR008947; PLC Nuclease.
DR Pfam; PF00882; Zn dep PLPC; 1.
DR PRINTS; PR00479; PRPHLPASEC.
DR PRODOM; PD003946; Zn dep PLPC; 1.
DR PROSITE; PS00384; PROKAR_ZN_DEPEND_PLPC; 1.
FT NON TER 275
SQ SEQUENCE 275 AA; 31213 MW; 87DA8752CF1BEC18 CRC64;

Query Match 31.7%; Score 482.5; DB 2; Length 275;
Best Local Similarity 36.9%; Pred. No. 1e-29;
Matches 97; Conservative 54; Mismatches 101; Indels 11; Gaps 3;

Qy 9 AAMVALAAPVQSVVFAQTNNSESPAPIL-RWSAEDKHNEGINSHLWTVNRAI 59
|||||
Db 14 ASIVALSQADEASACGDSVQDQAPHDIQKLPKLGWSAHPKSNINHLWLFNQAE 73
|||||
Qy 60 DIMSRNTTIVNPNETALINEWRADLENGIYSADYENPYDNTSYASHFYDPTGTITYPF 119
|||||
Db 74 KILAKHTVGAQLDLVRELKYNKIEAQIFDADHKNPYYDKNTFLSHFYNPKTKTYIAG 133
|||||
Qy 120 AKHAKETGAKYFNLAGQAYQNDQQOAPFYGLSLHYLGDVNPMPHAANFTNLSYPMGFH 179
|||||
Db 134 PPNKADTGTCTYFNISIEEYQDGNFEKAFYNLGLAIHYTIDISQPMHANNFTALSHPVGYH 193
|||||
Qy 180 SKYENFVDTIKNNYIVSDNSGYNWKGANPEDWIEGAAVAAKQDYPGVVNDTTKDFVKA 238
|||||
Db 194 CAYENYVDTFKQIFQASAESEAKW-FCYDDVSEPHENAKRAQADYKPVNTIINKSIYQ 252
|||||
Qy 239 AAVSQYADKRAEAVTPVTGKRL 261
|||||
Db 253 GLSDSQDRTWKAVRAATGKRL 275
|||||

RESULT 6

Q83TL2 PRELIMINARY; PRT; 242 AA.
ID Q83TL2;
AC Q83TL2;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Phospholipase (Fragment).
GN PLCB.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IMVW 1630;
RX MEDLINE=22535694; PubMed=12648840;
RA Schmid M., Walcher M., Bubert A., Wagner M., Wagner M.,
Schleifer K.-H.;
RT "Nucleic acid-based, cultivation-independent detection of Listeria
app. and genotypes of L. monocytogenes.";
RL FEMS Immunol. Med. Microbiol. 35:215-225(2003).
DR EMBL; AY150830; AAO19477.1; --
DR EMBL; AY150840; AAO19487.1; --
DR GO; GO:0004629; F:phospholipase C activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR008947; PLC Nuclease.
DR Pfam; PF00882; Zn dep PLPC; 1.
DR PRINTS; PR00479; PRPHLPASEC.
DR PRODOM; PD003946; Zn dep PLPC; 1.

DR PROSITE; PS00384; PROKAR_ZN_DEPEND_PLPC; 1.
FT NON TER 242 242
SQ SEQUENCE 242 AA; 27963 MW; CA45226804B99E87 CRC64;

Query Match 29.4%; Score 448; DB 2; Length 242;
Best Local Similarity 39.4%; Pred. No. 4.4e-27;
Matches 93; Conservative 45; Mismatches 86; Indels 12; Gaps 4;

QY 2 KKKVLAALAAVMA-LAAPVQSVV-----FAQT-----NNSESAP-IIRWSAEKHNREGIN 49
DB 4 KKKVGLMCLIASLVFPVPTIKANACCDEYLQTPAAPHDIDSKLPKLSWSADNPNTNDVN 63
QY 50 SHLWVRAIDMSRNTTIVNPETALLNEWRADLENGIYSADYENPYXNSTYASHFYD 109
DB 64 THYWLFKQAEKILAKOVNHRANLWELKFKDKQIAQGIYDADHKNPYDTSTFLSHFYN 123
QY 110 PDTGTYIIPFAKHAKEGTAKYFNLAGQAYQNDMQQAFYYLGLSLHYLGDVNPQMHAANF 169
DB 124 PDRONTYLPGFANAKITGAKYFNQSVTDYREGKFDTPAFYKLGIAIHVYTDISQPMHANF 183
QY 170 TNLSPMGFHSKYENFVDTIKNNYIVSDSGYNNWKGANPEDWIEGAVAAKQDYP 225
DB 184 TAISYPGPHYCAENYVDTIKHNYQATDMVAKRFGSDVDVKNLYENAKRAKRDYP 239

RESULT 7
Q84DK4 ID Q84DK4 PRELIMINARY; PRT; 242 AA.
AC Q84DK4
DT 01-JUN-2003 (TremBLrel. 24, Created)
DT 01-JUN-2003 (TremBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Phospholipase (Fragment).
GN PLCB.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22535694; PubMed=12648840;
RA Schmid M., Walcher M., Bubert A., Wagner M., Wagner M., Schleifer K.-H.;
RT "Nucleic acid-based, cultivation-independent detection of Listeria spp. and genotypes of L. monocytogenes.";
RL FEMS Immunol. Med. Microbiol. 35:215-225(2003).
DR EMBL; AY150833; AAO19480.1; -;
DR GO; GO:0004629; F:phospholipase C activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR008947; PLC_Nuclease.
DR Pfam; PF00882; Zn_dep_PLPC; 1.
DR PRINTS; PR00479; PRPHPLPASEC.
DR ProDom; PD003946; Zn_dep_PLPC; 1.
DR PROSITE; PS00384; PROKAR_ZN_DEPEND_PLPC; 1.
FT NON TER 242 242
SQ SEQUENCE 242 AA; 27890 MW; 7F3DEFD2AB43B372 CRC64;

Query Match 29.0%; Score 442; DB 2; Length 242;
Best Local Similarity 42.6%; Pred. No. 1.3e-26;
Matches 81; Conservative 38; Mismatches 71; Indels 0; Gaps 0;

QY 36 LRWSAEKHNREGINSHLWVRAIDMSRNTTIVNPETALLNEWRADLENGIYSADYEN 95
DB 50 LWSADNPNTDVTNHYLWFKQAEKILAKOVNHRANLWELKFKDKQIAQGIYDADHKN 109
QY 96 PYYDNTSYASHFYDPTGTYIIPFAKHAKEGTAKYFNLAGQAYQNDMQQAFYYLGLSLH 155
DB 110 PYYDNTSYASHFYDPTGTYIIPFAKHAKEGTAKYFNLAGQAYQNDMQQAFYYLGLSLH 169
QY 156 YLGDVNPQMHAANFTNLSPMGFHSKYENFVDTIKNNYIVSDSGYNNWKGANPEDWIEG 215
DB 170 YTTDISQPMHANFTAISYPGPHYCAENYVDTIKHNYQATDMVAKRFGSDVDVKNLYE 229
QY 216 AAVAAKQDYP 225
DB 230 NAKRAKRDYP 239

RESULT 9
Q84DK0 ID Q84DK0 PRELIMINARY; PRT; 242 AA.
AC Q84DK0
DT 01-JUN-2003 (TremBLrel. 24, Created)
DT 01-JUN-2003 (TremBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Phospholipase (Fragment).
GN PLCB.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.


```
Q84DK6
ID Q84DK6 PRELIMINARY; PRT; 236 AA.
AC Q84DK6
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Phospholipase (Fragment).
GN PLCB.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22535694; PubMed=12648840;
RA Schmid M., Walcher M., Bubert A., Wagner M.,
RA Schleifer K.-H.;
RT "Nucleic acid-based, cultivation-independent detection of Listeria
RT spp. and genotypes of L. monocytogenes.";
RL FEMS Immunol. Med. Microbiol. 35:215-225(2003).
DR EMBL, AY150833; AAO19479.1; -.
DR GO; GO:0004629; F:phospholipase C activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR008947; PLC Nuclease.
DR Pfam; PF00882; Zn dep_PLPC; 1.
DR ProDom; PD003946; Zn dep_PLPC; 1.
DR NON TER. 236
FT SEQUENCE 236 AA; 27083 MW; 4D6DC682E542897B CRC64;
SQ SEQUENCE 236 AA; 27083 MW; 4D6DC682E542897B CRC64;

Query Match 27.0%; Score 411; DB 2; Length 236;
Best Local Similarity 42.1%; Pred. No. 3.3e-24;
Matches 80; Conservative 36; Mismatches 68; Indels 6; Gaps 2;

Qy 36 LWSAEDKHNGINSHLVNRAIDMSRNTTIVNPNTALLNEWRADLENGIYSDYEN 95
Db 50 LWSADNPTNTDVTNTHYLFKQAEKILAKVDHMRANLNMELKNPKDKQIAQGIYDADHKN 109
Qy 96 PYDNSTYASHFYDPDGTGTYIPFAKHAKEGAKYFNLAGQAYQNDMQQAFYLGSLH 155
Db 110 PYDTSFLSHFYNDPDG-----FA-NAKITGARYFNQSVADYREGKEDTAFYXGLAIP 163
Qy 156 YLGVNQPMHAANFTNLSPMGFHSKYENFVDTIKNNYIVSDSNGYNNWKGANPEDWTIEG 215
Db 164 YTTDXQPMHANNFTAISYPRGYCHAYENYVDTIKHNTQATEDMVVKRFCSNDVKDMLYE 223
Qy 216 AAVAKQDYP 225
Db 224 NAKRAKADYP 233

RESULT 13
Q84DK2
ID Q84DK2 PRELIMINARY; PRT; 242 AA.
AC Q84DK2
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Phospholipase (Fragment).
GN PLCB.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22535694; PubMed=12648840;
RA Schmid M., Walcher M., Bubert A., Wagner M.,
RA Schleifer K.-H.;
RT "Nucleic acid-based, cultivation-independent detection of Listeria
RT spp. and genotypes of L. monocytogenes.";
RL FEMS Immunol. Med. Microbiol. 35:215-225(2003).
DR EMBL, AY150838; AAO19485.1; -.
DR GO; GO:0004629; F:phospholipase C activity; IEA.

Query Match 23.5%; Score 357.5; DB 2; Length 223;
Best Local Similarity 33.9%; Pred. No. 4.5e-20;
Matches 80; Conservative 42; Mismatches 83; Indels 31; Gaps 5;

Qy 2 KKKYLALAAVMA-LAAPVQSVV-----FAQT-----NNSESPAP-ILRWSAEDKHNGIN 49
Db 4 KKVVLGCLIASVLVFPVTTKANACCDBYLQTPAAPHDIDSKLPKLSWSADNPTNTDVN 63
Qy 50 SHLVNRAIDMSRNTTIVNPNTALLNEWRADLENGIYSDYENYNNSTYASHFYD 109
Db 64 THYLFKQAEKILAKVDHMRANLNMELKNPKDKQIAQGIYDADHKNPIYDTSTFLSHFYN 123

DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR008947; PLC Nuclease.
DR Pfam; PF00882; Zn dep_PLPC; 1.
DR ProDom; PD003946; Zn dep_PLPC; 1.
DR NON TER. 242
FT SEQUENCE 242 AA; 27840 MW; C916BE4DAAE71E8F CRC64;
SQ SEQUENCE 242 AA; 27840 MW; C916BE4DAAE71E8F CRC64;

Query Match 27.0%; Score 411; DB 2; Length 242;
Best Local Similarity 41.1%; Pred. No. 3.4e-24;
Matches 78; Conservative 35; Mismatches 77; Indels 0; Gaps 0;

Qy 36 LWSAEDKHNGINSHLVNRAIDMSRNTTIVNPNTALLNEWRADLENGIYSDYEN 95
Db 50 LWSADNPTNTDVTNTHYLFKQAEKILAKVDHMRANLNMELKNPKDKQIAQGIYDADHKN 109
Qy 96 PYDNSTYASHFYDPDGTGTYIPFAKHAKEGAKYFNLAGQAYQNDMQQAFYLGSLH 155
Db 110 PYDTSFLSHFYNDPDG-----FA-NAKITGARYFNQSVADYREGKEDTAFYXGLAIP 169
Qy 156 YLGVNQPMHAANFTNLSPMGFHSKYENFVDTIKNNYIVSDSNGYNNWKGANPEDWTIEG 215
Db 170 YTTDXQPMHANNFTAISYPRGYCHAYEXXVDTTIXHNTQATEDMVVKRFCSDDVKDMLYE 229
Qy 216 AAVAKQDYP 225
Db 230 NAKRAKADYP 239

RESULT 14
Q84DK7
ID Q84DK7 PRELIMINARY; PRT; 223 AA.
AC Q84DK7
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Phospholipase (Fragment).
GN PLCB.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NCTC 5105;
RX MEDLINE=22535694; PubMed=12648840;
RA Schmid M., Walcher M., Bubert A., Wagner M.,
RA Schleifer K.-H.;
RT "Nucleic acid-based, cultivation-independent detection of Listeria
RT spp. and genotypes of L. monocytogenes.";
RL FEMS Immunol. Med. Microbiol. 35:215-225(2003).
DR EMBL, AY150831; AAO19478.1; -.
DR GO; GO:0004629; F:phospholipase C activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR008947; PLC Nuclease.
DR Pfam; PF00882; Zn dep_PLPC; 1.
DR ProDom; PD003946; Zn dep_PLPC; 1.
DR NON TER. 223
FT SEQUENCE 223 AA; 25700 MW; AAD07392ACA37593 CRC64;
SQ SEQUENCE 223 AA; 25700 MW; AAD07392ACA37593 CRC64;

Query Match 23.5%; Score 357.5; DB 2; Length 223;
Best Local Similarity 33.9%; Pred. No. 4.5e-20;
Matches 80; Conservative 42; Mismatches 83; Indels 31; Gaps 5;

Qy 2 KKKYLALAAVMA-LAAPVQSVV-----FAQT-----NNSESPAP-ILRWSAEDKHNGIN 49
Db 4 KKVVLGCLIASVLVFPVTTKANACCDBYLQTPAAPHDIDSKLPKLSWSADNPTNTDVN 63
Qy 50 SHLVNRAIDMSRNTTIVNPNTALLNEWRADLENGIYSDYENYNNSTYASHFYD 109
Db 64 THYLFKQAEKILAKVDHMRANLNMELKNPKDKQIAQGIYDADHKNPIYDTSTFLSHFYN 123
```


This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2004, 14:32:07 ; Search time 13 Seconds
(without alignments)
1129.522 Million cell updates/sec

Title: PCT-US03-12556-2
Perfect score: 1522
Sequence: 1 MKKVLALAMVALAAPVQS.....EAQRTAGYIHLWFDYVNR 282

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1253.5	82.4	283	1 PHLD_BACCE	P33376 bacillus ce
2	1252.5	82.3	283	1 PHLC_BACCE	P09598 bacillus ce
3	511	33.6	289	1 PHLC_LISMO	P33378 listeria mo
4	254	16.7	398	1 PHLC_CLOBI	P20419 clostridium
5	244.5	16.1	399	1 PHLC_CLOSO	Q8VUZ6 clostridium
6	237	15.6	398	1 PLC2_CLOPE	Q9RF12 clostridium
7	229	15.0	398	1 PLC1_CLOPE	P15310 clostridium
8	213.5	14.0	399	1 PHLC_CLOHA	P59026 clostridium
9	205	13.5	398	1 PHLC_CLONO	Q46150 clostridium
10	102	6.7	562	1 EXG2_YEAST	P52911 saccharomyc
11	102	6.7	2358	1 MOXD_SCHPO	Q9Y719 schizosacch
12	97.5	6.4	363	1 OMPF_SALTI	Q56113 salmonella
13	97.5	6.4	363	1 OMPF_SALTY	P37432 salmonella
14	97	6.4	468	1 YBPM_ECOLI	P75733 escherichia
15	95.5	6.3	1283	1 OSH2_YEAST	Q12451 saccharomyc
16	95	6.2	624	1 YE70_METJA	Q58865 methanococc
17	94.5	6.2	418	1 HLT_VIBPA	Q99289 vibrio para
18	94	6.2	623	1 EXAA_PSEAE	Q9Z4J7 pseudomonas
19	94	6.2	837	1 XNVZ_CLOTM	P10478 clostridium
20	93.5	6.1	277	1 FMF4_ECOLI	P11900 escherichia
21	93.5	6.1	369	1 XYNA_PPRERU	P48789 prevotella
22	93.5	6.1	787	1 OXAA_CHLUM	Q9PK63 chlamydia m
23	93	6.1	333	1 XLNA_AGABI	O60206 agaricus bi
24	92.5	6.1	910	1 DNMJ_MYCPN	P75354 mycoplasma
25	92	6.0	782	1 Y044_UREPA	Q9P2A1 ureaplasma
26	91.5	6.0	474	1 PEDA_LACHE	Q48558 lactobacill
27	91	6.0	377	1 CCHL_SCHPO	O74794 schizosacch
28	90.5	5.9	257	1 OMPV_VIBAN	O08237 vibrio angu
29	90.5	5.9	787	1 OXAA_CHLTR	O84253 chlamydia t
30	90	5.9	423	1 P65_MYCPN	P53663 mycoplasma
31	90	5.9	1365	1 GTF5_STRDO	P29336 streptococ
32	89	5.8	235	1 NCAP_BUNSH	P03513 bunyavirus
33	89	5.8	695	1 CICA_PAEPP	P57092 paenibacill

ALIGNMENTS

RESULT 1

ID	PHLD_BACCE	STANDARD;	PRT;	283 AA.
AC	P33376;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Phospholipase C precursor (EC 3.1.4.3) (PLC) (Phosphatidylcholine cholinephosphohydrolase) (Cereolysin A).			
GN	CERA.			
OS	Bacillus cereus.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1396;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GP-4;			
RX	MEDLINE=8912149; PubMed=2536680;			
RA	Gilmore M.S., Cruz-Rodz A.L., Lelmeister-Waechter M., Kreft J., Goebel W.;			
RT	"A Bacillus cereus cytolytic determinant, cereolysin AB, which comprises the phospholipase C and sphingomyelinase genes: nucleotide sequence and genetic linkage."			
RL	J. Bacteriol. 171:744-753(1989).			
RN	[2]			
RP	SEQUENCE OF 50-142 FROM N.A.			
RA	Gilmore M.S., Gilmore K.S., Goebel W.;			
RT	"A new strategy for ordered DNA sequencing based on a novel method for the rapid purification of near-milligram quantities of a cloned restriction fragment."			
RL	Gene Anal. Tech. 2:108-114(1985).			
CC	- FUNCTION: Required, with sphingomyelinase to effect target cell lysis (hemolysis).			
CC	- CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = 1,2-diacylglycerol + choline phosphate.			
CC	- COFACTOR: Binds 3 zinc ions.			
CC	- SUBUNIT: Monomer.			
CC	- SIMILARITY: BELONGS TO THE BACTERIAL ZINC-METALLOPHOSPHOLIPASES C FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; M24149; AAA91819.1; -			
CC	EMBL; M35411; AAA22524.1; -			
CC	PIR; S18978; P50197.			
CC	HSSP; P09598; 1AH7.			
CC	InterPro; IPR008947; PLC Nuclease.			
CC	InterPro; IPR001531; Zn_dep_PLPC.			
CC	Pfam; PF00882; Zn_dep_PLPC_1.			
CC	PRINTS; PR00479; PRPHPLPASEC.			
CC	ProDom; PD003946; Zn_dep_PLPC; 1.			

PROSITE; PS00384; PROKAR_ZN_DEPEND_PLPC; 1. Hydrolase; Zinc; Signal; Zymogen; Hemolysis. POTENTIAL. 1 24 BY SIMILARITY. 25 38 PROPEP 39 283 CHAIN 39 283 METAL 39 39 METAL 52 52 METAL 93 93 METAL 107 107 METAL 156 156 METAL 160 160 METAL 166 166 METAL 180 180 METAL 184 184 SEQUENCE 283 AA; 32373 MW; 24D2E753402A6A44 CRC64;	Query Match Best Local Similarity 82.4%; Score 1253.5; DB 1; Length 283; Matches 226; Conservative 23; Mismatches 33; Indels 1; Gaps 1;
1 MKKVLALAAWVLAAPVQSVVFAQTNNSESAPIL-RNSAEDKHNEGINSHLWVNRAI 59	1 MKKVLALAAWVLAAPVQSVVFAQTNNSESAPIL-RNSAEDKHNEGINSHLWVNRAI 60
60 DIMSRNTTVNPNETALLNWRADLENGIYSADYENPYDNDSTYASHFYDDPTGTYIPF 119	60 DIMSRNTTVNPNETALLNWRADLENGIYSADYENPYDNDSTYASHFYDDPTGTYIPF 119
61 DIMSRNTTVLQKDRVALLNWRLENGIYAADYENPYDNTFFASHFYDDPDKGTYIPY 120	61 DIMSRNTTVLQKDRVALLNWRLENGIYAADYENPYDNTFFASHFYDDPDKGTYIPY 120
120 AKHAKETGAKYFNLAGOAYQYQNDMQQAFFYGLGLSHLYGLDVNQPMHAANFTNLSYPMGFH 179	120 AKHAKETGAKYFNLAGOAYQYQNDMQQAFFYGLGLSHLYGLDVNQPMHAANFTNLSYPMGFH 179
121 AKQAKETGAKYFKLAGESYKXNKKQAFYFGLGLSHLYGLDVNQPMHAANFTNLSYPOQGFH 180	121 AKQAKETGAKYFKLAGESYKXNKKQAFYFGLGLSHLYGLDVNQPMHAANFTNLSYPOQGFH 180
180 SKYENFVDTIKNNYIVSDSNGYNNWKCANPEDEMGAAVAAKODYPGVVDNTTKDWFVKA 239	180 SKYENFVDTIKNNYIVSDSNGYNNWKCANPEDEMGAAVAAKODYPGVVDNTTKDWFVKA 239
181 SKYENFVDVTIKNNYKVTGNGYNNWKTNPEDMHGAUVAKQDYAGIYNDNTKDFVRA 240	181 SKYENFVDVTIKNNYKVTGNGYNNWKTNPEDMHGAUVAKQDYAGIYNDNTKDFVRA 240
240 AVS0EYADKWRABVTPTVTKRLMEABQRTAGYIHLMFDTYVNR 282	240 AVS0EYADKWRABVTPTVTKRLMEABQRTAGYIHLMFDTYVNR 282
241 AVS0EYADKWRABVTPTVTKRLMDAQRVTAGYIQLMFDYTYNR 283	241 AVS0EYADKWRABVTPTVTKRLMDAQRVTAGYIQLMFDYTYNR 283

RESULT 2

ID	PHLC	BACCE	STANDARD;	PRT;	283 AA.
AC	R29538				
DT	01-MAR-1989	(Rel. 10, Created)			
DT	01-MAR-1989	(Rel. 10, Last sequence update)			
DT	28-FEB-2003	(Rel. 43, Last annotation update)			
DE	Phospholipase C precursor (EC 3.1.4.3) (PLC) (Phosphatidylcholine cholinephosphohydrolase) (Cereolysin A).				
GN	PLC.				
OS	Bacillus cereus.				
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.				
OX	NCBI_TaxID=1396;				
[1]					
SEQUENCE	FROM N.A.				
RC	STRAIN=SE-1;				
RC	MEDLINE=88313678; PubMed=3137122;				
RA	Johansen T., Holm T., Guddal P.H., Sletten K., Haugli F.B., Little C.;				
RT	"Cloning and sequencing of the gene encoding the phosphatidylcholine-				
RT	preferring phospholipase C of Bacillus cereus.";				
RL	Gene 65:293-304(1988).				
[2]					
SEQUENCE	FROM N.A.				
RC	STRAIN=VKM B-164;				
RC	MEDLINE=93249510; PubMed=8387306;				
RA	Kuzmin N.P., Gavrilenko I.V., Krukov V.M., Karpov A.V.;				
RT	"Nucleotide sequence of phospholipase C and sphingomyelinase genes				
RT	from Bacillus cereus BKM-B164 (letter).";				
RL	Bioorg. Khim. 19:133-138(1993).				
[3]					
SEQUENCE	OF 166-283 FROM N.A.				
RC	STRAIN=IAM 1208;				

```

FT TURN 96 103
FT STRAND 105 107
FT TURN 110 110
FT STRAND 111 114
FT TURN 115 115
FT STRAND 119 120
FT TURN 124 140
FT STRAND 141 142
FT TURN 144 161
FT STRAND 162 162
FT TURN 164 169
FT STRAND 172 173
FT TURN 177 178
FT STRAND 179 190
FT TURN 191 194
FT STRAND 203 204
FT TURN 210 223
FT STRAND 224 224
FT TURN 225 227
FT STRAND 228 228
FT TURN 231 240
FT STRAND 241 242
FT TURN 244 280
SQ SEQUENCE 283 AA; 32363 MW; ACS452EFP2E22B19 CRC64;

Query Match 82.3%; Score 1252.5; DB 1; Length 283;
Best Local Similarity 80.2%; Pred. No. 4.4e-92;
Matches 227; Conservative 23; Mismatches 32; Indels 1; Gaps 1;

Oy 1 MKKKVLALAAVVALAAPVQSVFAQTNNSSEAPIL-RWSAEDKNEGINSHLWVNR 59
Db 1 MKKKVLALAAVVALAAPVQSVFAQTNNSSEAPIL-RWSAEDKNEGINSHLWVNR 60

Oy 60 DIMSNTTIVNPNETALLNEWRADLENGIYSADYENPPYDNTYASHFYDPDTGTIYPF 119
Db 61 DIMSNTTIVKQDRVAQLNWRTELENGIYAADYENPPYDNTYASHFYDPDTGTIYPF 120

Oy 120 AKHAKETGAKYFNLAGQYQNDQMOQAFYGLSLHLYLGDVNPQMHAAFNLTSLYPMGFH 179
Db 121 AKQAKETGAKYFNLAGESYKNDKQAFYGLSLHLYLGDVNPQMHAAFNLTSLYPMGFH 180

Oy 180 SKYENFVDTIKNNYVSDSGVWNNKGANPEWIEGAAVAAKODYPGVVNDTKDWFVKA 239
Db 181 SKYENFVDTIKNNYVSDSGVWNNKGANPEWIEGAAVAAKODYPGVVNDTKDWFVKA 240

Oy 240 AVSQEYADKRAEVPVTKRLMEARVTAGYIHLWFDYVNR 282
Db 241 AVSQEYADKRAEVPVTKRLMDAQRVTAGYIQLWFDYVNR 283

RESULT 3
PHLC LISMO STANDARD; PRT; 289 AA.
ID PHLC LISMO STANDARD; PRT; 289 AA.
AC P33378.
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phospholipase C precursor (EC 3.1.4.3) (PLC) (Phosphatidylcholine
DE cholinephosphohydrolase) (Lecithinase).
GN PLCB OR PRTC OR LMO0205.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LO28 / Serovar 1/2c;
RX MEDLINE=92104678; PubMed=1309513;
RA Vazquez-Boland J.-A., Kocks C., Dramsi S., Ohayon H., Geoffroy C.,
RA Mengaud J., Cossart P.;
RT "Nucleotide sequence of the lecithinase operon of Listeria
RT monocytogenes and possible role of lecithinase in cell-to-cell
RT spread.";
RL Infect. Immun. 60:219-230(1992).

```

```

RN RP SEQUENCE FROM N.A.
RC STRAIN=EGD / Serovar 1/2a;
RX MEDLINE=92258410; PubMed=1582425;
RA Domann E., Wehlant J., Rohde M., Pistor S., Hartl M., Goebel W.,
RA Leimeister-Waechter M., Wuensther M., Chakraborty T.,
RT "A novel bacterial virulence gene in Listeria monocytogenes required
RT proline-rich region of vinculin.";
RL EMBO J. 11:1981-1990(1992).

RN RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero A., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunat F., Kurapkut G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Rammel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlant J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).

RN RP SEQUENCE OF 144-152.
RX MEDLINE=91267617; PubMed=1904842;
RA Geoffroy C., Raveneau J., Beretti J.L., Lecroisey A.,
RA Vazquez-Boland J.-A., Alouf J.E., Berche P.;
RT "Purification and characterization of an extracellular 29-kilodalton
RT phospholipase C from Listeria monocytogenes.";
RL Infect. Immun. 59:2382-2388(1991).

CC -!- FUNCTION: IMPORTANT ROLE IN THE INFECTIOUS PROCESS. MAY
CC CONTRIBUTE TO EFFICIENT LYSIS OF THE TWO-MEMBRANE VACUOLES THAT
CC SURROUND THE BACTERIA AFTER DIRECT CELL-TO-CELL SPREAD.
CC -!- CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = 1,2-
CC diacylglycerol + choline phosphate.
CC -!- COFACTOR: Binds 3 zinc ions (By similarity).
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL ZINC-METALLOPHOSPHOLIPASES C
CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; M82881; AAA25270.1; -
CC EMBL; X59723; CA442408.1; -
CC EMBL; AL591974; CAD00732.1; -
CC PIR; AF1100; AF1100.
CC PIR; S20888; S20888.
CC HSSP; P09598; 1A87.
CC Listlist; LMO00205; -
CC InterPro; IPR008947; PLC Nuclease.
CC InterPro; IPR001531; Zn_dep_PLPC.
CC Pfam; PF00882; Zn_dep_PLPC; 1.
CC PRINTS; PR00479; PRPHPLPASEC.
CC ProDom; PD003946; Zn_dep_PLPC; 1.
CC PROSITE; PS00384; PROKAR_ZN_DEPEND_PLPC; 1.
KW Hydroxylase; Zinc; Zymogen; Signal; Complete proteome.
FT SIGNAL 1 25
FT PROPEP 26 51
FT CHAIN 52 289
FT METAL 52 52
FT METAL 65 65
FT METAL 106 106
FT METAL 120 120

```



```
FT METAL 169 169 ZINC 1 (BY SIMILARITY).
FT METAL 173 173 ZINC 1 AND 3 (BY SIMILARITY).
FT METAL 179 179 ZINC 2 (BY SIMILARITY).
FT METAL 193 193 ZINC 2 (BY SIMILARITY).
FT METAL 197 197 ZINC 2 (BY SIMILARITY).
FT CONFLICT 5 5 K -> N (IN REF. 2).
FT CONFLICT 13 13 I -> T (IN REF. 2).
FT CONFLICT 222 222 D -> G (IN REF. 2).
SQ SEQUENCE 289 AA; 33277 MW; 0F35A2A3EDA6E372 CRC64;

Query Match 33.6%; Score 511; DB 1; Length 289;
Best Local Similarity 37.1%; Pred. NO. 2.2e-33;
Matches 106; Conservative 55; Mismatches 107; Indels 18; Gaps 5;

QY 2 KKKVLAAAMVA-LAAPQSVV-----PAQT-----NNSEGPAP-ILWASAEKHNKGIN 49
DB 4 KKKVLGMLIASLVLPVPTIRKANACCBYLOTFAAPHDIDSKLPKHSWADNPTNTDVN 63
QY 50 SHLWTVNRADIMSRNTTIVNPNTALLNWRADLENGIYADYENPYDNTSYASHFYD 109
DB 64 THYMLFKQAEKILAKDVNHRANLMLKKPKDKQIAQGIYDADHKNPYYDTSTFLSHFYN 123
QY 110 PDTGTYTIPFAKAKETGAKYFNLAGQAYQNDMQQAFYGLSLHYHLYGDNVQPMHAANF 169
DB 124 PDRDNTYLPGEANAKITGAKYFNQSVTDYREGKPDYAFYKGLAIHYTYDISQPMHANF 183
QY 170 TNLSPMGFHSKYENFVDTIKONYIVSDSNGYNNKGANPEDWIEGAVAQAQDYPGVVN 229
DB 184 TAISYPPGHCAYENYVDTIRHNYQATBDWAKRFCSDDVDKWLVENAKRAKADYPKLVN 243
QY 230 DTTKDPFKAAVQSYADKRAEVTPTGKRLMEAZQRTAGYIHLW 275
DB 244 AKTKSYLVG-----NSEWKDTVEPTGARLSDSQTLAGFLFW 283

RESULT 4
PHLC_CLOBI
ID PHLC_CLOBI STANDARD; PRT; 398 AA.
AC P20419, Q8VUZA; Q9S532;
AT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phospholipase C precursor (EC 3.1.4.3) (PLC) (Phosphatidylcholine
DE cholinophosphohydrolase) (Chp).
GN PLC OR CBP.
OS Clostridium bifementans.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1490;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=ATCC 638;
RX MEDLINE=89108588; PubMed=2536356;
RA Tao J.Y., Siebel C.;
RT "Cloning and expression of the phospholipase C gene from Clostridium
RT perfringens and Clostridium bifementans.";
RL Infect. Immun. 57:468-476(1989).
RN [2]
RP SEQUENCE FROM N.A., AND CONSTRUCTION OF A HYBRID ENZYME.
RC STRAIN=ATCC 638;
RX MEDLINE=99307183; PubMed=10377104;
RA Jepson M., Howelle A.M., Bullifent H.L., Bolgiano B., Crane D.T.,
RA Miller J., Holley J., Jayasekera P., Titball R.W.;
RT "Differences in the carboxy-terminal (putative phospholipid binding)
RT domains of Clostridium perfringens and Clostridium bifementans
RT phospholipases C influence the hemolytic and lethal properties of
RT these enzymes.";
RL Infect. Immun. 67:3297-3301(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ 1012;
RA Karsawa T., Wang X., Maegawa T., Michiwa Y., Miwa K., Nakamura S.;
RT "C. sordellii phospholipase C.";
```

```
Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
[4]
RN REVIEW
RX MEDLINE=20465460; PubMed=11008117;
RA Jepson M., Titball R.W.;
RT "Structure and function of clostridial phospholipases C.";
RL Microbes Infect. 2:1277-1284(2000).
CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
CC cell membranes and cause cell rupture. Binds to eukaryotic
CC 10-fold less activity towards sphingomyelin than its C.perfringens
CC counterpart, is approximately 100-fold less hemolytic against
CC mouse erythrocytes and at least 100-fold less toxic in mice.
CC -!- CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = 1,2-
CC diacylglycerol + choline phosphate.
CC -!- COPACTOR: Binds 3 zinc and 3 calcium ions (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- DOMAIN: The protein is composed of 2 domains; the N-terminal
CC domain contains the phospholipase C active site (PLC), in a cleft
CC which is also occupied by the 3 zinc ions. The C-terminal domain
CC is a putative phospholipid-recognition domain, which shows
CC structural homology with phospholipid-binding C2-like domains from
CC a range of eukaryotic proteins. The ability to bind membrane
CC phospholipids in a Ca(2+) dependent manner is conferred by this C-
CC terminal domain (By similarity).
CC -!- MISCELLANEOUS: A hybrid protein between the N-terminus of
CC C.bifermentans and the C-terminus of C.perfringens has an activity
CC intermediate between the two.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL ZINC-METALLOPHOSPHOLIPASES C
CC FAMILY.
CC -!- SIMILARITY: Contains 1 PLAT domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF072123; AAD41623.1; -.
DR EMBL; AB061869; BAB83265.1; -.
DR PIR; B30565; B30565.
DR InterPro; IPR001024; Lipoxigenase_LH2.
DR InterPro; IPR008976; PLAT_LH2.
DR InterPro; IPR008947; PLC_Nuclease.
DR InterPro; IPR001531; Zn_dep_PLPC.
DR Pfam; PF01477; PLAT; 1.
DR Pfam; PF00882; Zn_dep_PLPC; 1.
DR PRINTS; PR00479; PRPHPLPASEC.
DR PRODOM; PD003946; Zn_dep_PLPC; 1.
DR PROSITE; PS50095; PLAT_1.
DR PROSITE; PS00384; PROKAR_ZN_DEPEND_PLPC; 1.
KW Hydrolase; Toxin; Hemolysis; Calcium; Calcium-binding; Zinc;
KW Signal.
FT SIGNAL 1 26 BY SIMILARITY.
FT CHAIN 27 398 PHOSPHOLIPASE C.
FT DOMAIN 29 272 PHOSPHOLIPASE C.
FT DOMAIN 273 281 LINKER.
FT DOMAIN 282 398 PLAT.
FT METAL 27 27 ZINC 1 (BY SIMILARITY).
FT METAL 37 37 ZINC 1 (BY SIMILARITY).
FT METAL 82 82 ZINC 3 (BY SIMILARITY).
FT METAL 94 94 ZINC 3 (BY SIMILARITY).
FT METAL 152 152 ZINC 3 (BY SIMILARITY).
FT METAL 156 156 ZINC 1 AND 3 (BY SIMILARITY).
FT METAL 162 162 ZINC 2 (BY SIMILARITY).
FT METAL 174 174 ZINC 2 (BY SIMILARITY).
FT METAL 178 178 ZINC 2 (BY SIMILARITY).
FT METAL 297 297 CALCIUM 1 (VIA CARBONYL OXYGEN) (BY
FT SIMILARITY).
FT METAL 298 298 CALCIUM 3 (BY SIMILARITY).
FT METAL 299 299 CALCIUM 3 (BY SIMILARITY).
```


QY 93 YENPYDINSTYASHFYDPTGTYYI-----PFAKHAKETGAKYFNLAGQAYQNDM 143
 Db 85 YDPNAYD--LFQDHFDPDTGNFLLDNKWTYASPIYDTATQVTKFATLAEWKKGN 142
 QY 144 QAAFFYGLSLHYLDGVNQPMAHNFNLSPMGPHSKYENFVDTIKNNYIVSDSNGYWN 203
 Db 143 KEATFLGQGHYLDGLNTPYHAANVTAVDSP--GHVKYETFVENKXENYALNTAG--- 196
 QY 204 WGANPEWIEGAJAAVAAKQDPGVNDTKDFWFKAAVSQYVADKWRLEVFTVTKRLME 263
 Db 197 -----NDTTOGIY-KDAVANKDFQW-----MTQNSVKY 224
 QY 264 AQRVTAGYI-----HLWFD 277
 Db 225 AKKAKALYISHSTMKHND 244

RESULT 6

PLC2_CLOPE
 ID PLC2_CLOPE STANDARD; PRT; 398 AA.
 AC Q9RF12,
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Phospholipase C precursor (EC 3.1.4.3) (PLC) (Phosphatidylcholine
 cholinephosphohydrolase) (Alpha-toxin) (Hemolysin) (Lecithinase).
 GN PLC OR CPA.
 OS Clostridium perfringens.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1502;
 RN [1]
 RP SEQUENCE FROM N.A., CHARACTERIZATION, AND X-RAY CRYSTALLOGRAPHY (2.4
 ANGSTROMS) OF THE OPEN FORM.
 RC STRAIN=SWCP;
 RX MEDLINE=22005774; PubMed=12009886;
 RA Justin N., Walker N., Bullifant H.L., Songer G., Bueschel D.M.,
 Jost H., Naylor C.E., Miller J., Moes D.S., Tibball R.W., Basak A.K.,
 "The first strain of Clostridium perfringens isolated from an avian
 RT source has an alpha-toxin with divergent structural and kinetic
 RT properties.";
 RL Biochemistry 41:6253-6262 (2002).
 RN [2]
 RP REVIEW.
 RA Tibball R.W., Naylor C.E., Basak A.K.;
 RT "The Clostridium perfringens alpha-toxin.";
 RL Anaerobe 5:51-64 (1999).
 RN [3]
 RP REVIEW.
 RX MEDLINE=20465460; PubMed=11008117;
 RA Japeon M., Tibball R.W.;
 RT "Structure and function of clostridial phospholipases C.";
 RL Microbes Infect. 2:1277-1284 (2000).
 CC -1- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
 cell membranes and cause cell rupture. Constitutes an essential
 CC virulence factor in gas gangrene. Binds to eukaryotic membranes
 CC where it hydrolyzes both phosphatidylcholine and sphingomyelin,
 CC causing cell rupture. The diacylglycerol produced can activate
 CC both the arachidonic acid pathway, leading to modulation of the
 CC inflammatory response cascade and thrombosis, and protein kinase
 CC C, leading to activation of eukaryotic phospholipases and further
 CC membrane damage.
 CC -1- CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = 1,2-
 CC diacylglycerol + choline phosphate.
 CC -1- COFACTOR: Binds 3 zinc and 3 calcium ions.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- DOMAIN: The protein is composed of 2 domains; the N-terminal
 CC domain contains the phospholipase C active site (PLC), in a cleft
 CC which is also occupied by the 3 zinc ions. The C-terminal domain
 CC is a putative phospholipid-recognition domain, which shows
 CC structural homology with phospholipid-binding C2-like domains from
 CC a range of eukaryotic proteins. The ability to bind membrane
 CC phospholipids in a Ca(2+) dependent manner and toxicity is

CC conferred by this C-terminal domain, which also contributes to the
 CC sphingomyelinase activity.
 CC -1- MISCELLANEOUS: This bacteria was isolated from a diseased swan.
 CC -1- SIMILARITY: Contains 1 PLAT domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF204209; AAF20094.1; -.
 DR PDB; 1KHO; 19-JUN-02.
 DR InterPro; IPR001024; Lipoxigenase_LH2.
 DR InterPro; IPR008976; PLAT_LH2.
 DR InterPro; IPR008947; PLC_Nuclease.
 DR InterPro; IPR001531; Zn Dep_PLPC.
 DR Pfam; PF00882; Zn dep PLPC; 1.
 DR PRINTS; PD00479; BRPHEPLPASEC.
 DR PRODOM; PD003945; Zn dep PLPC; 1.
 DR PROSITE; PS50095; PLAT; 1.
 DR PROSITE; PS00384; PROKAR_ZN_DEPEND_PLPC; FALSE NEG.
 DR HydroLase; Toxin; Hemolysis; Calcium; Calcium-Binding; Zinc;
 KW Signal; 3D-structure.
 FT SIGNAL 1 28
 FT CHAIN 29 398
 FT DOMAIN 29 274
 FT DOMAIN 275 283
 FT DOMAIN 284 398
 FT METAL 29 29
 FT METAL 39 39
 FT METAL 84 84
 FT METAL 96 96
 FT METAL 154 154
 FT METAL 158 158
 FT METAL 164 164
 FT METAL 176 176
 FT METAL 180 180
 FT METAL 299 299
 FT METAL 300 300
 FT METAL 301 301
 FT METAL 321 321
 FT METAL 322 322
 FT METAL 324 324
 FT METAL 325 325
 FT METAL 326 326
 FT METAL 326 326
 FT METAL 365 365
 FT SEQUENCE 398 AA; 45598 MW; EF3C95AF603CD5A7 CRC64;
 SQ
 Query Match 15.6%; Score 237; DB 1; Length 398;
 Best Local Similarity 26.4%; Pred. No. 1.6e-11;
 Matches 79; Conservative 47; Mismatches 113; Indels 60; Gaps 13;
 QY 1 MKKKYLLAALAAVVALAAPVQSVVFAQTNNSESPAPILRWSAEDKHNEGINSHLWIVNRAID 60
 Db 1 MKRKIKLLICATITATSWAV-----RTRKYVAM---DGRADGTGTHAMIAIQGT 48
 QY 61 I-----MSRNTTIVNPNETALLNWRADLENGIYSADYENPYDINSTYASHFYDPTGT 116
 Db 49 ILENDLSSNEPEVIRNLEILKQNMHDLQLGSTPYDYDKNAYD--LYQDHFWDPTDNNF 106
 QY 117 IPFAK-----HAKETGAKYFNLAGQAYQNDMQQAFVGLSLHYLDGVNQPMAH 167
 Db 107 TDKSKWYLSYISPTDAESQIRKFSALARYWKRGYKQATFYLGAMHYFGDADTPYHAA 166
 QY 168 NPTNLSYPMGFHSKYENFVDTIKNNYIVSDSNGYWNKGANPEDWIEGAJAAVAAKQDPGV 227
 Db 167 NVTAVDSP--GHVKETFAEDRDKQYKI-----NTTGSKTNDAFY-SNLTNEDF--- 213

STRAIN=13 / Type A, K2211 / Type A, L9 / Type D, NCIB 10663 / Type D
NCIB 10691 / Type B, NCIB 10748 / Type E, and PB6K / Type A;
MEDLINE=96099301; PubMed=8522524;
Tetsutsui K., Minami J., Matsushita O., Katayama S., Taniguchi Y.,
Nakamura S., Nishioka M., Okabe A.;
"Phylogenetic analysis of phospholipase C genes from Clostridium
perfringens types A to E and Clostridium novyi.";
J. Bacteriol. 177:7164-7170(1995).

[8]
SEQUENCE FROM N.A., AND USE AS A VACCINE.
STRAIN=13 / Type A, CER 89L43 / Type A, CER 89L1105 / Type A, and
CER 89L1216 / Type A;
MEDLINE=96146062; PubMed=8581165;
Gincer A., Williamson E.D., Dessy F., Coppe P., Bullifant H.,
Howells A.M., Titball R.W.;
"Molecular variation between the alpha-toxins from the type strain
(NCTC 8237) and clinical isolates of Clostridium perfringens
associated with disease in man and animals.";
Microbiology 142:191-198(1996).

[9]
SEQUENCE FROM N.A.
STRAIN=K21340 / Type A;
MEDLINE=96336195; PubMed=8709860;
Kameyama K., Matsushita O., Katayama S.-I., Minami J., Maeda M.,
Nakamura S., Okabe A.;
"Analysis of the phospholipase C gene of Clostridium perfringens
K21340 isolated from Antarctic soil.";
Microbiol. Immunol. 40:255-263(1996).

[10]
SEQUENCE FROM N.A.
STRAIN=13 / Type A;
MEDLINE=21664373; PubMed=11792842;
Shimizu T., Ohnari K., Hirakawa H., Ohshima K., Yamashita A.,
Shiba T., Ogasawara N., Hattori H., Kuhara S., Hayashi H.;
"Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater.";
Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).

[11]
MUTAGENESIS OF HISTIDINE RESIDUES.
STRAIN=NCTC 8237 / ATCC 13124 / CN 1491 / Type A;
MEDLINE=95173092; PubMed=7868589;
Nagahama M., Okagawa Y., Nakayama T., Nishioka E., Sakurai J.;
"Site-directed mutagenesis of histidine residues in Clostridium
perfringens alpha-toxin.";
J. Bacteriol. 177:1179-1185(1995).

[12]
MUTAGENESIS OF ZINC-BINDING, SUBSTRATE-BINDING, AND CATALYTIC
RESIDUES.
STRAIN=8-6 / Type A;
MEDLINE=96294749; PubMed=8698464;
Guillouard I., Garnier T., Cole S.T.;
"Use of site-directed mutagenesis to probe structure-function
relationships of alpha-toxin from Clostridium perfringens.";
Infect. Immun. 64:2440-2444(1996).

[13]
MUTAGENESIS OF THR-300, AND CONSERVED ASPARTATE AND TYROSINE RESIDUES.
STRAIN=8-6 / Type A;
MEDLINE=98086094; PubMed=9426125;
Guillouard I., Alzari P.M., Saitou B., Cole S.T.;
"The carboxy-terminal C2-like domain of the alpha-toxin from
Clostridium perfringens mediates calcium-dependent membrane
recognition.";
Mol. Microbiol. 26:867-876(1997).

[14]
MUTAGENESIS OF ASP-297; TYR-303; TYR-335; TYR-359 AND ASP-364.
STRAIN=NCTC 8237 / ATCC 13124 / CN 1491 / Type A;
MEDLINE=20389543; PubMed=10931204;
Alape-Giron A., Flores-Diaz M., Guillouard I., Naylor C.E.,
Titball R.W., Rucavado A., Lomonte B., Basak A.K., Gutierrez J.M.,
Cole S.T., Thelestam M.;
"Identification of residues critical for toxicity in Clostridium
perfringens phospholipase C, the key toxin in gas gangrene.";
Eur. J. Biochem. 267:5191-5197(2000).

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; AF525415; AA089377.1; --
DR InterPro; IPR001024; Lipoxigenase_LH2.
DR InterPro; IPR008976; PLAT_LH2.
DR InterPro; IPR008947; PLC_Nuclease.
DR InterPro; IPR001531; Zn_dep_PLPC.
DR Pfam; PF00882; Zn_dep_PLPC; 1.
DR PRINTS; PR00479; PRPHPLPASEC.
DR PRODOM; PD003946; Zn_dep_PLPC; 1.
DR PROSITE; PS00995; PLAT; 1.
DR PROSITE; PS00384; PROKAR_ZN_DEPEND_PLPC; FALSE NEG.
DR Hemolysis; Hydrolase; Calcium; Calcium-binding; Zinc; Toxin;
KW Signal.
FT SIGNAL 1 28 BY SIMILARITY.
FT CHAIN 29 399 PHOSPHOLIPASE C.
FT DOMAIN 29 274 PHOSPHOLIPASE C.
FT DOMAIN 275 282 LINKER.
FT DOMAIN 283 399 PLAT.
FT METAL 28 28 ZINC 1 (BY SIMILARITY).
FT METAL 38 38 ZINC 1 (BY SIMILARITY).
FT METAL 83 83 ZINC 3 (BY SIMILARITY).
FT METAL 95 95 ZINC 3 (BY SIMILARITY).
FT METAL 153 153 ZINC 3 (BY SIMILARITY).
FT METAL 157 157 ZINC 1 AND 3 (BY SIMILARITY).
FT METAL 163 163 ZINC 2 (BY SIMILARITY).
FT METAL 175 175 ZINC 2 (BY SIMILARITY).
FT METAL 179 179 ZINC 2 (BY SIMILARITY).
FT METAL 228 228 CALCIUM 1 (VIA CARBONYL OXYGEN) (BY
FT METAL 299 299 CALCIUM 3 (BY SIMILARITY).
FT METAL 300 300 CALCIUM 3 (BY SIMILARITY).
FT METAL 320 320 CALCIUM 2 (BY SIMILARITY).
FT METAL 321 321 CALCIUM 2 (BY SIMILARITY).
FT METAL 323 323 CALCIUM 2 (VIA CARBONYL OXYGEN) (BY
FT METAL 324 324 CALCIUM 3 (BY SIMILARITY).
FT METAL 325 325 CALCIUM 2 (BY SIMILARITY).
FT METAL 325 325 CALCIUM 3 (BY SIMILARITY).
FT METAL 363 363 CALCIUM 1 (BY SIMILARITY).
SQ SEQUENCE 399 AA; 45942 MW; 97A8FA5537083BEA CRC64;
Query Match 14.0%; Score 213.5; DB 1; Length 399;
Best Local Similarity 25.9%; Pred. No. 1.1e-09;
Matches 76; Conservative 40; Mismatches 108; Indels 69; Gaps 14;
2 KKKVLALAAWVALAAPVQSVVFAQTNNSESPAPILRWSAEDKHNEGINSHLWIVNRAIDI 61
3 KKKILKFCISAVLS-----FTLPSGYKSYA----W----DGKVDGTGTHALIVTQAVEI 48
62 MSRTNTIIVNP-----NETALLNEWRADLENGIYSADEPNYDNTSYASHFYDPDGTGTYI 117
49 LKNDVISTSPLSVKENFKILESNLKQLQSGTYPDYDKAY--ALYQDFWDPDPTDNN-- 104
118 PPAKHAK-----ETG-----AKYENLAGAQYQNDWQQAFFYGLSLHYLGDVNPQMH 165
105 -FTKDSKMYLAVGINETGESQURKULFALAKDSEKKNYEQATWLLGQGLHYFGDPTTPH 163
166 AANFTNLSYPMGFHSHKYENFVDTIKNNYIVSDNSGYNWKNGANPEDWIEGAAVAQDYP 225
164 PSNVT--AVDSAGHKFYFYEGKDSYKLFHA-----GAN-----SVKEFPY 204
226 GVNDDTKD-WFVKAAVSGEYADKKRAEVTPTVTKRLMEARQRTAGYIHLWFD 277
205 TTLQNTLNDNWT-----EYSGRW-----AKKAKNMYAHATMSHSWKD 243

RESULT 9
PHLC_CLONO

ID AC 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phospholipase C precursor (EC 3.1.4.3) (PLC) (Phosphatidylcholine
DE cholinephosphohydrolase) (gamma-toxin).
GN PLC.
OS Clostridium novyi.
OS Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1542;
RN [1]_TaxID=1542;
RP SEQUENCE FROM N.A.
RX STRAIN=CL49 / Type A;
RX STRAIN=96099301; PubMed=8522524;
RA Tetsui K., Minami J., Matsushita O., Katayama S.-I., Taniguchi Y.,
RA Nakamura S., Nishioka M., Okabe A.;
RT "Phylogenetic analysis of phospholipase C genes from Clostridium
RT perfringens types A to E and Clostridium novyi.",
RL J. Bacteriol. 177:7164-7170(1995).
RN [2]
RP CHARACTERIZATION.
RC STRAIN=IID 140 / Type A;
RX MEDLINE=76040159; PubMed=241423;
RA Taguchi R., Ikezawa H.;
RT "Phospholipase C from Clostridium novyi type A. I.",
RL Biochim. Biophys. Acta 409:75-85(1975).
RN [3]
RP REVIEW.
RX MEDLINE=20465460; PubMed=11008117;
RA Jepson M., Titball R.W.;
RT "Structure and function of clostridial phospholipases C.",
RL Microbes Infect. 2:1277-1284(2000).
CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
CC cell membranes and cause cell rupture. Binds to eukaryotic
CC membranes where it hydrolyzes phosphatidylcholine, sphingomyelin
CC and phosphatidylethanolamine. The diacylglycerol produced can
CC activate both the arachidonic acid pathway, leading to modulation
CC of the inflammatory response cascade and thrombosis, and protein
CC kinase C, leading to activation of eukaryotic phospholipases and
CC further membrane damage (By similarity). This enzyme is hemolytic
CC against horse erythrocytes.
CC -!- CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = 1,2-
CC diacylglycerol + choline phosphate.
CC -!- COFACTOR: Binds 3 zinc and 3 calcium ions (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- DOMAIN: The protein is composed of 2 domains; the N-terminal
CC domain contains the phospholipase C active site (PLC), in a cleft
CC which is also occupied by the 3 zinc ions. The C-terminal domain
CC is a putative phospholipid-recognition domain, which shows
CC structural homology with phospholipid-binding C2-like domains from
CC a range of eukaryotic proteins. The ability to bind membrane
CC phospholipids in a Ca(2+) dependent manner and toxicity is
CC conferred by this C-terminal domain, which also contributes to the
CC sphingomyelinase activity.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL ZINC-METALLOPHOSPHOLIPASES C
CC FAMILY.
CC -!- SIMILARITY: Contains 1 PLAT domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; D31215; BAA06851.1; --
DR InterPro; IPR001024; Lipoxigenase_LH2.
DR InterPro; IPR008976; PLAT_LH2.
DR InterPro; IPR008947; PLC_Nuclease.
DR InterPro; IPR001531; Zn_dep_PLPC.


```
Db 314 -----HYFNLTGANYSSQDI-----LYVDH-----HYEVFTD 340
QY 172 LSPYMGFHSKYENFV-----DTIKNN-----IVSDNGYWNWKGANPE--DWIEGAARAAK 221
Db 341 AQLAETQFARIENIYNGDSIHKELSFHVAUGE-----WSGALTDCATWLNQVGVGAR 394
QY 222 QD--YGVVNDTTFKQFVKAASQEVADKRAEVPVTKRLMEARQV 268
Db 395 YDGSYNTTFTNDKPVGTCISQNSLADWTQDYDRV--RQPIEAQLAT 442

RESULT 11
MOKD SCHPO
ID MOKD SCHPO STANDARD; PRT; 2358 AA.
AC Q9Y7I9; O94638;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cell wall alpha-1,3-glucan synthase mok13 (EC 2.4.1.183).
GN MOK13 OR SPBC16D10.05.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Katayama S., Dai H., Arellano M., Perez P., Toda T.;
RT "Fission yeast alpha-glucan synthase Moki localizes closely with actin
RT and play a role essential for cell morphogenesis and protein kinase C
RT function.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
EX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitz E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallard J.C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomycetes pombe.";
RL Nature 415:871-880(2002).
CC -!- CATALYTIC ACTIVITY: UDP-glucose + {alpha-D-glucosyl-(1,3)} (N) =
CC UDP + {alpha-D-glucosyl-(1,3)} (N+1).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
```

```
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; AB018382; BAA76559.1; -.
DR EMBL; AL035637; CAB38509.1; -.
DR PIR; T39569; T39569.
DR PIR; T43432; T43432.
DR GenDB_SPombe; SPBC16D10.05; -.
DR InterPro; IPR006047; Alpha_amy_cat.
DR Pfam; PF00128; alpha-amyase; 1.
KW Cell wall; Transferase; Glycosyltransferase.
FT CONFLICT 120 120 V -> VRRVMLCSLTNKV (IN REF. 1).
SQ SEQUENCE 2358 AA; 269192 MW; A175577C9D8AD731 CRC64;

Query Match 6.7%; Score 102; DB 1; Length 2358;
Best Local Similarity 18.4%; Pred. No. 6.9;
Matches 67; Conservative 43; Mismatches 115; Indels 140; Gaps 15;

QY 3 KKVLAALAAWALAAPVQSVVFAQ-----TNNSSPAPILRWSAEDKHNEGINSHLMTV 55
Db 4 KNILVLNLLISIPRLVFTAKYDERESLWNLNQNSATDPLDYWGKWNHQYHPSDDMQV 63
QY 56 NRAIDMSRNTTIVNPNETALINWRADELENGIYSADYENPYDNTYASHFYDPTGTT 115
Db 64 -----PFTVILDKWK-----DGDPRNEANTIVEYDIY----- 93
QY 116 YIPFAKHAKETGAKYFNLAGQAYQNDQMOQAFYFGLSLHYLGD---VNQPMHAANFTNL 172
Db 94 -----ETG---FRNGDGIIGLGLSLDYLEIMGIKVIYIAGTPFLNQPMGADQYSPL 141
QY 173 SYPM-----GFHSKYENFVDIT----- 189
Db 142 DYTLDHHSHTGAQWRDTIEIHRGFGYLVLDLTISTIGDLIGFRKLYNLSTTFFSLFEHE 201
QY 190 ---KNNYIVSDSNGYWNW-----KGANPEDWIE-GAAVAAKQDYPGVND----- 230
Db 202 AVKSNVIYPD---WNFTNKYDPKCELPFRWGEDGAPWVI--DIVGCVSDSDFOYGDTE 255
QY 231 ---TTKOW---FVKAASVQEVADKRAEVT-----PVTGRLMEARQV 270
Db 256 AFGTHPDERQLSKFASVQDRLEWRPVSSEKLKHFACMIITAMLDVDGFRIDKATQITVD 315
QY 271 YIHLW 275
Db 316 FLASW 320

RESULT 12
ID OMPF_SALTI STANDARD; PRT; 363 AA.
AC Q56113;
DT 15-DEC-1998 (Rel. 37, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Outer membrane protein F precursor (Porin ompF) (Outer membrane
DE protein S3).
GN OMPF OR OMP53 OR STY1002 OR T1935.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM5S-1;
RA Fernandez-Mora M., Calva E.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
```

RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrell B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TY2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 RT and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -!- FUNCTION: OMPF is a porin that forms passive diffusion pores which
 CC allow small molecular weight hydrophilic materials across the
 CC outer membrane. It is also a receptor for the bacteriophage T2 (By
 CC similarity).
 CC -!- SUBUNIT: Homotrimer (By similarity).
 CC -!- SIMILARITY: Belongs to the Gram-negative porin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC
 CC EMBL; X89577; CA61905.1; -;
 DR EMBL; AL627268; CAD05399.1; -;
 DR EMBL; AE016840; AAO69550.1; -;
 DR HSP; P02931; 1GFN.
 DR InterPro; IPR003229; OMP 2.
 DR InterPro; IPR001702; Porin Gram-ve.
 DR Pfam; PF00267; Gram-ve_porins; 1.
 DR PRINTS; PR00182; ECOLNEIPORIN.
 DR ProDom; PD000808; OMP 2; 1.
 DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.
 KW Outer membrane; Transmembrane; Porin; Phage recognition; Signal;
 KW Complete proteome.
 FT SIGNAL 1 22 BY SIMILARITY.
 FT CHAIN 23 363 OUTER MEMBRANE PROTEIN P.
 FT TRANSMEM 23 363 BY SIMILARITY.
 FT DOMAIN 29 29 PERIPLASMIC (BY SIMILARITY).
 FT TRANSMEM 30 45 BY SIMILARITY.
 FT DOMAIN 46 56 EXTRACELLULAR (BY SIMILARITY).
 FT TRANSMEM 57 69 BY SIMILARITY.
 FT DOMAIN 70 71 PERIPLASMIC (BY SIMILARITY).
 FT TRANSMEM 72 84 BY SIMILARITY.
 FT DOMAIN 85 99 EXTRACELLULAR (BY SIMILARITY).
 FT TRANSMEM 100 108 BY SIMILARITY.
 FT DOMAIN 109 109 PERIPLASMIC (BY SIMILARITY).
 FT TRANSMEM 110 117 BY SIMILARITY.
 FT DOMAIN 118 154 EXTRACELLULAR (BY SIMILARITY).
 FT TRANSMEM 155 161 BY SIMILARITY.
 FT DOMAIN 162 169 PERIPLASMIC (BY SIMILARITY).
 FT TRANSMEM 170 181 BY SIMILARITY.
 FT DOMAIN 182 192 EXTRACELLULAR (BY SIMILARITY).
 FT TRANSMEM 193 203 BY SIMILARITY.
 FT DOMAIN 204 204 PERIPLASMIC (BY SIMILARITY).
 FT TRANSMEM 205 217 BY SIMILARITY.
 FT DOMAIN 218 230 EXTRACELLULAR (BY SIMILARITY).
 FT TRANSMEM 231 242 BY SIMILARITY.
 FT DOMAIN 243 243 PERIPLASMIC (BY SIMILARITY).
 FT TRANSMEM 244 256 BY SIMILARITY.
 FT DOMAIN 257 272 EXTRACELLULAR (BY SIMILARITY).
 FT TRANSMEM 273 285 BY SIMILARITY.
 FT DOMAIN 286 287 PERIPLASMIC (BY SIMILARITY).
 FT TRANSMEM 288 301 BY SIMILARITY.

FT DOMAIN 302 312 EXTRACELLULAR (BY SIMILARITY).
 FT TRANSMEM 313 324 BY SIMILARITY.
 FT DOMAIN 325 326 PERIPLASMIC (BY SIMILARITY).
 FT TRANSMEM 327 336 BY SIMILARITY.
 FT DOMAIN 337 353 EXTRACELLULAR (BY SIMILARITY).
 FT TRANSMEM 354 363 BY SIMILARITY.
 FT CONFLICT 307 307 D -> G (IN REF. 1).
 SQ SEQUENCE 363 AA; 40106 MW; P5059B37EA516859 CRC64;

Query Match 6.4%; Score 97.5; DB 1; Length 363;

Best Local Similarity 23.5%; Pred. No. 1.5;
 Matches 67; Conservative 28; Mismatches 109; Indels 81; Gaps 17;

QY 1 MKKVLALAAVLAAPVQSVFPAQTNNSSPAPILRWSAEDKINSHLMIV----- 55
 DB 2 MKRILAAVPAALAAATANA--AEIYNKDG-----NKLDLGKAVGRHWTTGDSK 52
 QY 56 ---NRAIDIMSRNTTIVNPETALLNEW---RADLENG-----LY 89
 DB 53 NADQTYAQIGKGETQINTDLTG-FGQWEVTKADRAEGEQQNSNLVRLAPAGLKYAEVG 111
 QY 90 SADYENPY---YDNSTY---ASHFYDPTGTITYPPAKHAKETGAKYENLAGQAYQNDM 143
 DB 112 SIDYGRNYGIYDYVESYDWPYFSGETWGTAYDNTYMTSRAGL-----LTRYNSD- 163
 QY 144 QQAFYYL--GLS--LHYLGDVNPQMAANFTN---LSYPM-----GFHSKYENFVDTI 189
 DB 164 ---FFGLVDGLSFGIQGK-NQDNHNSINQNGQGVGYTMAYEFGVGTAAYSNKR- 218
 QY 190 KNNYIVSDSGYNNWKGANPDMTEGAATAAKQDYPGVVNDTKD 234
 DB 219 -NQQDRDGN-----GDRAESWAVGAKYDANNVLAAYVAETRN 256

RESULT 13

ID OMPF SALTY STANDARD; PRT; 363 AA.
 AC P37432;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Outer membrane protein F precursor (Porin ompf) (Outer membrane
 DE protein 1A) (Outer membrane protein 1A) (Outer membrane protein B).
 GN OMPF OR STM0999.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2;
 RA Venegas A., Gomez I., Bruce E., Martinez M.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 EX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).
 CC -!- FUNCTION: OMPF IS A PORIN THAT FORMS PASSIVE DIFFUSION PORES WHICH
 CC ALLOW SMALL MOLECULAR WEIGHT HYDROPHILIC MATERIALS ACROSS THE
 CC OUTER MEMBRANE. IT IS ALSO RECEPTOR FOR THE BACTERIOPHAGE T2 (BY
 CC similarity).
 CC -!- SUBUNIT: Homotrimer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
 CC (By similarity).
 CC -!- SIMILARITY: Belongs to the Gram-negative porin family.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Z31594; CA83471.1; -;
DR EMBL; AE008743; AAL19933.1; -;
DR PIR; S43159; S43159.
DR HSP; P02931; IGFN.
DR StyGene; SG10264; OmpF.
DR InterPro; IPR003229; OMP 2.
DR Pfam; PF00267; Gram-ve_porins; 1.
DR PRINTS; P00182; ECOLNEIPORIN.
DR ProDom; PD000808; OMP 2; 1.
DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Phage recognition; Signal;
KW Complete proteome.

FT SIGNAL 1 22 BY SIMILARITY.
FT CHAIN 23 363 OUTER MEMBRANE PROTEIN F.
FT TRANSMEM 23 28 BY SIMILARITY.
FT DOMAIN 29 29 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 30 45 BY SIMILARITY.
FT DOMAIN 46 56 EXTRACELLULAR (BY SIMILARITY).
FT TRANSMEM 57 69 BY SIMILARITY.
FT DOMAIN 70 71 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 72 84 BY SIMILARITY.
FT DOMAIN 85 99 EXTRACELLULAR (BY SIMILARITY).
FT TRANSMEM 100 108 BY SIMILARITY.
FT DOMAIN 109 109 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 110 117 BY SIMILARITY.
FT DOMAIN 118 154 EXTRACELLULAR (BY SIMILARITY).
FT TRANSMEM 155 161 BY SIMILARITY.
FT DOMAIN 162 169 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 170 181 BY SIMILARITY.
FT DOMAIN 182 192 EXTRACELLULAR (BY SIMILARITY).
FT TRANSMEM 193 203 BY SIMILARITY.
FT DOMAIN 204 204 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 205 217 BY SIMILARITY.
FT DOMAIN 218 230 EXTRACELLULAR (BY SIMILARITY).
FT TRANSMEM 231 242 BY SIMILARITY.
FT DOMAIN 243 243 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 244 256 BY SIMILARITY.
FT DOMAIN 257 272 EXTRACELLULAR (BY SIMILARITY).
FT TRANSMEM 273 285 BY SIMILARITY.
FT DOMAIN 286 287 PERIPLASMIC (BY SIMILARITY).
FT TRANSMEM 288 301 EXTRACELLULAR (BY SIMILARITY).
FT DOMAIN 302 312 BY SIMILARITY.
FT TRANSMEM 313 324 PERIPLASMIC (BY SIMILARITY).
FT DOMAIN 325 326 BY SIMILARITY.
FT TRANSMEM 327 336 EXTRACELLULAR (BY SIMILARITY).
FT DOMAIN 337 353 BY SIMILARITY.
FT TRANSMEM 354 363 E -> A (IN REF. 1).
FT CONFLICT 232 232 T -> V (IN REF. 1).
FT CONFLICT 350 360
SQ SEQUENCE 363 AA; 40048 MW; F6069B34E9516859 CRC64;

Query Match 6.4%; Score 97.5; DB 1; Length 363;
Best Local Similarity 23.5%; Pred. No. 1.5;
Matches 67; Conservative 28; Mismatches 109; Indels 81; Gaps 17;

QY 1 MKKKVLALAAWVALAAPVQSVVFAQTNSPAPILRWSEAPDKHNGSHLVIV----- 55
DB 2 MKRKILAAVAPILAAATANA--AEIYNKDG-----NKLGLYKAVGRHWTTTGDGSK 52

QY 56 ---NRAIDMSRTTIVNPETALLNEW-----RADLENG-----IY 89
DB 53 NADQTYAQIGPKGETQINTDLTG-FQWEYRTKADRAEGEQNSNLVRLAFAGLKYAEVG 111

QY 90 SADYENPY---YDNSTY---ASHFYVDPDTGTYIPFAKHAKETGAKYFNLAGQAYONQDM 143
DB 112 SIDYGRNYGIVYDVESYTDMAFYSGETWGGAYTDNYMTSAGGL-----LTYRNSD- 163
QY 144 QQAQFYL--GLS--LHYLGDVNPQMPHAANFTN---LSYPM-----GFHSKYENFVDTI 189
DB 164 ---FPGLVDGLSFGIYQYQK-NQDNHLSINSONGDVGVTMAYEPDGFVGVTAYSNSKRT- 218
QY 190 KNNYIVSDNSNGYWNWKGANPEDWIEGAAVAARQDYPGVVNDTTKD 234
DB 219 -NQDQRDGN-----GDRAESSWAGAKYDANNVYLAAYVAETRN 256

RESULT 14

YBFM_ECOLI
ID YBFM_ECOLI STANDARD; PRT; 468 AA.
AC P75733;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ybFM.
GN YBFM OR B0681.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.,
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayaishi K., Honjo A.,
RA Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AE000172; AAC73775.1; -;
DR EMBL; D90707; BAA35329.1; ALT_INIT.
DR PIR; H64802; H64802.
DR EcoGene; EG13659; ybFM.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 468 AA; 52780 MW; 954B5A778AE1C2EA CRC64;

Query Match 6.4%; Score 97; DB 1; Length 468;
Best Local Similarity 22.0%; Pred. No. 2.3;
Matches 78; Conservative 46; Mismatches 136; Indels 94; Gaps 20;

QY 2 KKKVLALA-----AWVALAAPVQSVVFAQTNSPAPILRWSEAPDKHNG 47
DB 7 KRSLTALAAGVTAMSGFNPMPPEARAEFGIDDTLTGGIYYWQRDRKQVTDGDKYKTN 66

```

QY 48 INSHLWIVNR-----AIDI-----MSRNTTIIVNPETALL-----NEWRA 82
Db 67 LSHSTWNLDFQSGYAADMFGLDIAAFTALEMAENGDSHPNIAFSKSNKAYDEWSG 126
QY 83 DLENGI--YSADYENPYDNTYASHFYDDPTGTT-----YIPFAKHAKETGAKY-FN 132
Db 127 D-KSGISYKAAAKPKY--GPWBARAGYIQGTQTLAPHSFPGTVOGAEGANFDYG 183
QY 133 LAGO---AYQNDQQQAFYIYGLSHYLGVDVNPQMHAAFNLSYPMGFHSKYENFVDTI 189
Db 184 DAGALSFYSMTNNEYKAPWHELMDEFYQND-----KTKVDYLYHSFGAKYD-----F 230
QY 190 KNNYIV-----SDSNGYWN--WKGANPEDWIEGAAVAQDYPGV-----VNDTTKD- 234
Db 231 KNNFVLEAFAQGAEGYIDQYFAKASYKFDIAGSPITTSYQYGTGRKVDKDRSVDNLLYDGT 290
QY 235 -WFKAAVSQEYAD--KWRAEYTPV--TGKRLMEARQVTAGY-----IHLWFD 277
Db 291 AWLQALTFGYRAADVVDLLEGTWYKADGQQGYFLQRMTPTYASSNGRLDIWMD 344

RESULT 15
OSH2 YEAST
ID OSH2 YEAST STANDARD; PRT; 1283 AA.
AC Q12451.P89891;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Oysterol-binding protein homolog 2.
GN OSH2 OR YDL019C OR D2845.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RA Urrestarazu L.A., Andre B., Vissers S.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBSJ databases.
RN [2]
RP SUBCELLULAR LOCATION.
RX MEDLINE=21301806; PubMed=11408574;
RA Levine T.P., Munro S.;
RT "Dual targeting of Oshlp, a yeast homologue of oysterol-binding
protein, to both the Golgi and the nucleus-vacuole junction.";
RL Mol. Biol. Cell 12:1633-1644 (2001).
RN [3]
RP GENETIC ANALYSIS.
RX MEDLINE=21135676; PubMed=11238399;
RA Beh C.T., Cool L., Phillips J., Rine J.;
RT "Overlapping functions of the yeast oysterol-binding protein
homologues.";
RL Genetics 157:1117-1140 (2001).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic. When bound to oysterols it
translocate to the periphery of Golgi membranes.
CC -1- SIMILARITY: Belongs to the OSBP family.
CC -1- SIMILARITY: Contains 2 ANK repeats.
CC -1- SIMILARITY: Contains 1 PH domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DB EMBL; Z74066; CAA98577.1; -
DB EMBL; Z74067; CAA98578.1; -
DB EMBL; Z48432; CAA88340.1; -
DB PIR; S52500; S52500.
DB GeneOnline; 140261; -
DB SGD; S0002177; OSH2.

```

```

DR GO; GO:0005935; C:bud neck; IDA.
DR GO; GO:0005886; C:plasma membrane; IDA.
DR GO; GO:0006594; P:steroid biosynthesis; IGI.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000648; Oysterol_BP.
DR InterPro; IPR001849; PH.
DR Pfam; PF00203; ank; 2.
DR Pfam; PF01237; Oysterol_BP; 1.
DR Pfam; PF00169; PH; 1.
DR SMART; SM00248; ANK; 2.
DR SMART; SM00233; PH; 1.
DR PROSITE; PS50088; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 2.
DR PROSITE; PS01013; OSBP; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW Lipid transport; Transport; ANK repeat; Repeat; Golgi stack.
FT REPEAT 106 134
FT REPEAT 206 235
FT DOMAIN 289 386
SQ SEQUENCE 1283 AA; 145795 MW; D521957460E7F7C3 CRC64;

Query Match 6.3%; Score 95.5; DB 1; Length 1283;
Best Local Similarity 25.9%; Pred. No. 11;
Matches 72; Conservative 31; Mismatches 104; Indels 71; Gaps 17;

QY 1 MKKKVLALAAWALAAPVQSVVPAQTNNSESPAPILRWSAEDKHNEGINSH----- 51
Db 838 VKETKEIASVIGEKTIIVAVTTVQKKEY---LLKEGSLVGVGDIKKRLSMDKDRP 894
QY 52 ---LWIVNRAI--DIMSRTTIIVNPNE-TALLNEWRADLENGIYSADYENPYDNTSYAS 105
Db 895 KISLWAVLKSMVGKDMTRMTLPVTFNEPTSLQORVAEDLE-----YSELLDQAA 943
QY 106 HFYDPDGTGYI-----PEAKHAKETGAKYFN-LAQQA--YQNDQQQAFYIYGLSHY 156
Db 944 TFDSTLRTLYAAFTASSYASTTKRV-AKFPNPLGTEFTEYSRDPQYRFFTEQVSHH- 1001
QY 157 LGDVNQPMHAANFTNLSYPMGFHSKYENFVDTTKNNYIVSDSN-GYWN----- 203
Db 1002 -----PPISATWTE-SPRWDFWG--ESFVDTKFGSRFNVKHLGLHKLKRPNDNEKEE 1052
QY 204 ---WKGANPEDWIEGAAVAQD-DYPG---VNDTTKD 234
Db 1053 LYTWK--KPNNTVIGILIGNPQVDNHGCVNVNHTTGD 1088

Search completed: July 27, 2004, 14:46:31
Job time : 15 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2004, 14:31:12 ; Search time 54 Seconds
(without alignments)
1475.526 Million cell updates/sec

Title: PCT-US03-12556-2
Perfect score: 1522
Sequence: 1 MKKKVLAAMVALAAPVQS.....EAQRTAGYIHLWFTYVNR 282

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	511	33.6	289	5	Abb47676 Listeria
2	220	14.5	370	2	Aar52563 Clostridi
3	108.5	7.1	2140	6	Abu01020 S. pneumo
4	108.5	7.1	2140	6	Abu45746 Protein e
5	100	6.6	444	4	Abg08390 Novel hum
6	98	6.4	414	4	Aag80920 Plasidida
7	96.5	6.3	836	5	Aag68288 Modified
8	96.5	6.3	836	6	Abu08984 B. latero
9	96.5	6.3	871	5	Aag68285 Brevibaci
10	96.5	6.3	871	5	Abu08981 B. latero
11	95.5	6.3	1283	6	Abr52701 Protein s
12	94.5	6.2	192	3	Aag52327 Arabidops
13	94.5	6.2	192	3	Aag05009 Arabidops
14	94.5	6.2	305	3	AAG05008 Arabidops
15	94.5	6.2	305	3	AAG52326 Arabidops
16	94.5	6.2	305	5	Abb90956 Herbicida
17	94	6.2	192	2	Aay86159 S. pneumo
18	94	6.2	288	3	Aag10306 Arabidops
19	94	6.2	288	3	Aag24900 Arabidops
20	94	6.2	288	3	AAG50140 Arabidops
21	94	6.2	294	3	AAG24899 Arabidops
22	94	6.2	294	3	AAG50139 Arabidops
23	94	6.2	294	3	AAG10305 Arabidops
24	94	6.2	323	3	AAG24898 Arabidops
25	94	6.2	324	3	AAG50138 Arabidops

26	94	6.2	346	3	AAG10304 Arabidops
27	94	6.2	399	2	Aaw93146 C. thermo
28	94	6.2	837	3	Aay70519 Clostridi
29	93.5	6.1	479	5	ABG30886 H. pylori
30	93.5	6.1	787	6	ABU27268 Protein e
31	93	6.1	370	6	ABM72521 Staphyloc
32	93	6.1	664	4	Aab48342 S. pneumo
33	93	6.1	2120	3	Aay81710 Streptoco
34	92.5	6.1	304	2	Aaw34599 Amaranthu
35	92.5	6.1	304	2	Aaw85564 Seed stor
36	92.5	6.1	304	2	Aaw83352 Amaranthu
37	92	6.0	1984	4	ABG01338 Novel hum
38	91.5	6.0	238	6	ABM67131 Photorhab
39	90.5	5.9	165	3	Aag05010 Arabidops
40	90.5	5.9	165	3	AAG52328 Arabidops
41	90.5	5.9	296	3	AAG50135 Arabidops
42	90.5	5.9	296	3	AAG05829 Arabidops
43	90.5	5.9	449	5	ABP65417 Bifidobac
44	90.5	5.9	584	6	ABU19765 Protein e
45	90.5	5.9	1320	5	ABG60631 Human pot

ALIGNMENTS

RESULT 1
ABB47676
ID ABB47676 standard; protein; 289 AA.
XX
AC ABB47676;
XX
DT 05-FEB-2002 (first entry)
XX
DE Listeria monocytogenes protein #380.
XX
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.
XX
OS Listeria monocytogenes.
XX
PN WO200177335-A2.
XX
PD 18-OCT-2001.
XX
PF 11-APR-2001; 2001WO-FR001118.
XX
PR 11-APR-2000; 2000FR-00004629.
XX (INSP) INST PASTEUR.
PA Buchrieser C, Frangeul L, Couve E, Rusniok C, Psihi H, Dehoux P;
PI Dussauget O, Chetouani F, Nedjati H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Krest J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX
WPI; 2002-010914/01.
DR Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and related
PT polypeptides.
XX Claim 6; SEQ ID NO 381; 192pp; French.

The present invention relates to the genome sequence of Listeria monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of it are useful for selecting probes and primers for detecting genes in L. monocytogenes and related organisms, and for studying genetic polymorphisms and other genomes. The present sequence is a protein encoded by the genome sequence of the present invention. Proteins

expressed from the genome sequence are useful for raising specific antibodies, identification of L. monocytogenes and related organisms, and for biosynthesis and biodegradation, especially biosynthesis of Vitamin B12. The genome sequence and proteins encoded by it are also useful for selecting compounds that regulate gene expression and cell replication and modulate L. monocytogenes-related diseases. In addition, the genome sequence and proteins encoded by it are useful in pharmaceutical and vaccines compositions for the treatment or prevention of infections by L. monocytogenes and related organisms. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 289 AA;

Query Match 33.6%; Score 511; DB 5; Length 289;
Best Local Similarity 37.1%; Pred. No. 5.7e-39;
Matches 106; Conservative 55; Mismatches 107; Indels 18; Gaps 5;

Qy 2 KKVJLALAMVA-LAAPVQSVV-----FAQT-----NNSESPAP-ILWMSAEDKHNEGIN 49
Db 4 KKVVLGMLIASVLVFPVTIKANACCDEYLOTPAAPHDIDSKLPHKLSWSADNPTNTDVN 63

Qy 50 SHLWIVNRAIDIMSRNTTIVNPNETALLNEWRADLENGIYADYENPYDNTSYASHPYD 109
Db 64 THYWLFKQAEKLAKDVNHRANLMLNKKPKDKIAQGIYDADHKNPYDTSFLSHFYN 123

Qy 110 PDTGTYTTPFAKHAKETGAKYFNLAGQAYQNQDMQQAFFYGLSLHYLGDVNOPMHAANP 169
Db 124 PDRDNTYLPGFANAKITGAKYFNQSVTDYREGKFDTAFFYKGLAIHYTYTDSQPMHANF 183

Qy 170 TNLSPMGPHSKYENFVDTIKNNYIVSDSNGYNNKGANPEDWIEGAAVAAKQDYPGVN 229
Db 184 TAISYPPGYHCAYENYVDTIKNNYQATQEDMVAKRFCSDDDVDKWLNYENAKRAKADIPKLVN 243

Qy 230 DTTKDMFVKAASQEVADKWRAEVPVTGKRLMEARQVTAQYIHLW 275
Db 244 ATKXSYLVG-----NSEWKKDVTVEPTGARLDSQQTLAGFLFW 283

RESULT 2
ID AARS2563 standard; protein; 370 AA.
XX AARS2563;
XX AARS2563;
DT 25-MAR-2003 (revised)
DT 30-MAY-1994 (first entry)
XX Clostridium perfringens alpha-toxin.
XX Alpha-toxin; AT; gas gangrene; vaccine; bacterium; Cpa; conjugate;
KW phospholipase C; sphingomyelin-hydrolysis.
XX Clostridium perfringens.
XX WO9323543-A1.
XX
XX 25-NOV-1993.
XX 20-MAY-1993; 93WO-GB001039.
XX 20-MAY-1992; 92GB-00010717.
XX 23-JUL-1992; 92GB-00015655.
XX (MINA) UK SEC FOR DEFENCE.
XX
XX Titball RW, Williamson ED;
XX
XX WPI; 1993-386575/48.
XX N-PSDB; AAQ62305.
XX New peptides from C-terminus of Clostridium perfringens alpha toxin -

PT useful in protective vaccines against gas gangrene, also derived antibodies for passive immunisation.
XX
XX Disclosure; Page 16-17; 29pp; English.
XX
XX Peptides, or conjugates, have the amino acid sequence of the epitopes of C. perfringens alpha-toxin (AT) from amino acids 261-300 but lack C. phospholipase C and/or sphingomyelin-hydrolyzing activity. The peptides induce a protective immune response against AT when admin. to human or animals. (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 370 AA;

Query Match 14.5%; Score 220; DB 2; Length 370;
Best Local Similarity 26.5%; Pred. No. 1.1e-11;
Matches 68; Conservative 40; Mismatches 95; Indels 54; Gaps 10;

Qy 46 EGINSHLWIVNRAIDIMSRNTTIVNPNET-----ALLNEWRADLENGIYSADYENPYDINS 101
Db 6 DGTGTHAMIVTQGVSILENDLSKNEPESVRKNLEILKENHMLQLGSGTYPDYDKNAYD-- 63

Qy 102 TYASHFYDPTGTGYTP-----FAKHAKETG-----AKYFNLAGQAYQNQDMQQAFFYGL 152
Db 64 LYQDFWDPDPTDNNFSDKNSWYLAISIPDTGSEQIRKFSALARYEWQRGNYKQATFVLGE 123

Qy 153 SLHYLGDVNOPMHAANFNLISYPMGPHSKYENFVDTIKNNYIVSDSNGYNN----- 203
Db 124 AMHYFGDIDTPYHPANVT--AVDSAGHVKEPTFAERKEQYKINTAGCKTNEAFYTDILK 181

Qy 204 -----WGANPEDWIEGAAVAAKQDYPGVNDDTKDMFVKAASQEVAD-KWRAEVPV 256
Db 182 NKDFNMS-----KEYARGPARKTSIY-----YSHASHMSHSDDDWYAAKVT-- 224

Qy 257 TGKRLMEARQVTAQYI 273
Db 225 ----LANSQKGTAGYII 237

RESULT 3
ID ABU01020 standard; protein; 2140 AA.
XX ABU01020;
XX AC
XX AC
DT 23-OCT-2003 (revised)
DT 11-FEB-2003 (first entry)
XX
XX S. pneumoniae type 4 strain protein from coding region #590.
XX Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
KW gene therapy; vaccine.
XX Streptococcus pneumoniae; type 4 strain.
XX OS
XX WO200277021-A2.
XX
XX 03-OCT-2002.
XX
XX 27-MAR-2002; 2002WO-IB002163.
XX
XX 27-MAR-2001; 2001GB-00007658.
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
XX Masignani V, Tettelin H, Fraser C;
XX
XX WPI; 2003-040579/03.
XX N-PSDB; ABX06302.
XX
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae, useful as medicaments for treating or preventing a disease or infection

PT useful in protective vaccines against gas gangrene, also derived antibodies for passive immunisation.
XX
XX Disclosure; Page 16-17; 29pp; English.
XX
XX Peptides, or conjugates, have the amino acid sequence of the epitopes of C. perfringens alpha-toxin (AT) from amino acids 261-300 but lack C. phospholipase C and/or sphingomyelin-hydrolyzing activity. The peptides induce a protective immune response against AT when admin. to human or animals. (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 370 AA;

Query Match 14.5%; Score 220; DB 2; Length 370;
Best Local Similarity 26.5%; Pred. No. 1.1e-11;
Matches 68; Conservative 40; Mismatches 95; Indels 54; Gaps 10;

Qy 46 EGINSHLWIVNRAIDIMSRNTTIVNPNET-----ALLNEWRADLENGIYSADYENPYDINS 101
Db 6 DGTGTHAMIVTQGVSILENDLSKNEPESVRKNLEILKENHMLQLGSGTYPDYDKNAYD-- 63

Qy 102 TYASHFYDPTGTGYTP-----FAKHAKETG-----AKYFNLAGQAYQNQDMQQAFFYGL 152
Db 64 LYQDFWDPDPTDNNFSDKNSWYLAISIPDTGSEQIRKFSALARYEWQRGNYKQATFVLGE 123

Qy 153 SLHYLGDVNOPMHAANFNLISYPMGPHSKYENFVDTIKNNYIVSDSNGYNN----- 203
Db 124 AMHYFGDIDTPYHPANVT--AVDSAGHVKEPTFAERKEQYKINTAGCKTNEAFYTDILK 181

Qy 204 -----WGANPEDWIEGAAVAAKQDYPGVNDDTKDMFVKAASQEVAD-KWRAEVPV 256
Db 182 NKDFNMS-----KEYARGPARKTSIY-----YSHASHMSHSDDDWYAAKVT-- 224

Qy 257 TGKRLMEARQVTAQYI 273
Db 225 ----LANSQKGTAGYII 237

RESULT 3
ID ABU01020 standard; protein; 2140 AA.
XX ABU01020;
XX AC
XX AC
DT 23-OCT-2003 (revised)
DT 11-FEB-2003 (first entry)
XX
XX S. pneumoniae type 4 strain protein from coding region #590.
XX Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
KW gene therapy; vaccine.
XX Streptococcus pneumoniae; type 4 strain.
XX OS
XX WO200277021-A2.
XX
XX 03-OCT-2002.
XX
XX 27-MAR-2002; 2002WO-IB002163.
XX
XX 27-MAR-2001; 2001GB-00007658.
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
XX Masignani V, Tettelin H, Fraser C;
XX
XX WPI; 2003-040579/03.
XX N-PSDB; ABX06302.
XX
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae, useful as medicaments for treating or preventing a disease or infection

PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
PT ear infection.

XX

Claim 1: SEO ID NO 1180: 56pp: English.

The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the specification (available on a computer readable format), or its fragment, expressed from 2469 of 2489 identified DNA coding regions from the *Streptococcus pneumoniae* type 4 strain genomic sequence appearing as A855454. Also included are an antibody which binds one of the proteins, treating a patient by administering the protein, DNA or antibody (in a composition), a kit comprising first and second primers, which are the nucleic acid cited above or fragments between nucleotides 8-100 of a sequence not defined in the specification, for amplifying a target sequence contained within a *Streptococcus nucleic acid* sequence, where the first primer is substantially complementary to the target sequence and the second primer is substantially complementary to the complement of the target sequence, and where the parts of the primers having substantial complementarity define the termini of the target sequence to be amplified, assay comprising contacting a test compound with the protein, and determining whether the test compound binds to the protein and a *Streptococcus pneumoniae* bacterium, where one or more genes encoding the proteins has been rendered inactive. The proteins, nucleic acid molecules, antibody and compositions are useful as medicaments for treating or preventing a disease or infection due to *Streptococcus* bacteria, particularly *S. pneumoniae*, such as pneumonia, sepsis, otitis media or ear infection. They are also useful in developing vaccines, diagnostics and antibiotics. The methods are useful for identifying immunodominant proteins. The present sequence is one of the 2469 proteins expressed by the identified coding regions from the genomic sequence. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences). (Updated on 23-OCT-2003 to standardise OS field)

Db 1 MKKSTVLSITTAAVILAAVAPNEVVLADTSSSEDALNIDSKEKVAENKEHENIHSAMET 60
Qy 53 -----WIVNRADIMSRNTTIVN--PNETALLNEWRADLENGIYADY-----ENP- 96
Db 61 SQDPFKKXTAVIKEXEVVSKNPFVIDNNTSNEEAKIKEENSNNKSGQDYDTSFVKNKTENPK 120
Qy 97 YDNTSYASHFVDPDTGTTTYPFAXHAKETGAKY-----ENLAGOAYO-----NQD----- 142
Db 121 KEDKVYIAEFKXESGGEKAKELSSLNKTKVLYTYDRIFN--GSAIETTPDNLDKIKQI 178
Qy 143 -----MQQAFYVLGL--SLHYLGDVNPQMHAANF-----TNLSYPMGF 178
Db 179 EGISVERAQAQVQPMWNNHARKEIGVEEALDYLSKINAPF-GKNFDRGWMVISNIDTGTDY 237
Qy 179 HSKYENFVDTIK-----NNYIVSD-----SNGYWNWKGANPEWIEGAIVAA 220
Db 238 RHKAMRIDDDAKASMRFKEDLKGTDKNYWLSDKIPHAFNYNGGKITVEKYDDG----- 292
Qy 221 KQDY-----PGVV--NDTTKD 234
Db 293 -RDYFDPHGMHAGILAGNDTEQD 315

RESULT 5

ABG08390
ID ABG08390 standard; protein; 444 AA.

XX ABG08390;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #8381.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

PD 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR N-P8DB; AAS72577.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

PS Claim 20; SEQ ID NO 38749; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 444 AA;

Query Match 6.6%; Score 100; DB 4; Length 444;

Best Local Similarity 22.3%; Pred. No. 2.1;

Matches 48; Conservative 36; Mismatches 79; Indels 52; Gaps 9;

Qy 2 KKKVLALAAVMA-----LAAPVQSVVFAQTNNSESPAPILRWSAEDKHNEGINSHLWIVNR 57

Db 255 KSSILDMGRLISSVENVPKYEGETESSSVSNVSSPIKLM-FLSEVKSSGKVYTLTSGV- 312

Qy 58 AIDIMSRNTTIVNPN-----TALLNEWRADLENGIYADYENPYDNTSYASHFDPDPTGT 114

Db 313 -----TSHSNVLPSEKPTTHVTEETKTETNEDISNANSEN-----YHSSHYDPTDFQ 360

Qy 115 TYIPPAKHAKETGAKYFNLAGQAYQNQDMQQAFFVLGLSLHYLGDVN-----QPMHAANFT 170

Db 361 RELNKFHAKETA-----GSSTFIFGIDNSDKFQEPKDNSS 397

Qy 171 NLSYPMGFHSHKYENFVDTIKNIYIVSDSNGYWNWK 205

Db 398 SATDP-SFNKKTR---XTKKNSRSPSCETDXATNWK 428

RESULT 6

AAG80920

ID AAG80920 standard; protein; 414 AA.

XX AAG80920;

DT 28-AUG-2001 (first entry)

DE Plastidial delta 9 ACP desaturase partial clone protein sequence.

XX Moss; Physcomitrella patens; lipid metabolism related protein; LMRP;
KW lipid biosynthesis; lipid modification; lipid degradation; cofactor;
KW fatty acid transport; genetic engineering; fatty acid; enzyme; plant;
KW microorganism; polyunsaturated fatty acid; oilseed plant; maize; wheat;
KW biotic stress tolerance; abiotic stress tolerance; rye; oat; triticale;
KW rice; barley; soybean; peanut; cotton; rapeseed; canola; manihot; pepper;
KW sunflower; tagetes; potato; tobacco; eggplant; tomato; vicia; pea;
KW alfalfa; coffee; cacao; tea; Salix; oil palm; coconut; perennial grass;
KW forage crop.

OS Physcomitrella patens.

XX WO200138484-A2.

PD 31-MAY-2001.

XX 22-NOV-2000; 2000WO-EP011615.

XX 25-NOV-1999; 99WO-EP009108.

PA (BADI) BASF PLANT SCI GMBH.

PI Lerchl J, Renz A, Ehrhardt T, Reindl A, Cirpus P, Bischoff F;

FR Frank M, Freund A, Duwenig E, Schmidt R, Reski R;

DR WPI; 2001-367669/38.

XX Nucleic acids encoding lipid metabolism related proteins from

PT Physcomitrella patens useful to produce fine chemicals in modified

PD 26-DEC-2002.
 XX 17-MAY-2001; 2001US-00858525.
 XX 17-MAY-2001; 2001US-00858525.
 XX (BOET/) BOETS A.
 PA (ARNA/) ARNAUT G.
 PA (RIEJ/) RIE J V.
 PA (DAMM/) DAMME N.
 XX
 XX Boets A, Arnaut G, Rie JV, Damme N;
 PI WPI; 2003-361872/34.
 DR N-PSDB; ABX94353.
 XX
 XX New insecticidal secreted proteins and DNA sequences, useful in making
 PT insecticidal compositions or transgenic plants to protect plants from
 PT damage caused by insects, particularly coleopteran insects.
 XX
 XX Claim 4; Page 32-34; 38pp; English.
 XX
 XX The invention relates to a new insecticidal secreted protein comprising:
 CC (a) amino acid sequence of the smallest active toxin of the protein
 CC comprising a sequence appearing as ABU08981 (ISP1A) or ABU08982 (ISP2A);
 CC or (b) the amino acid sequence of the protease-digestion fragment of the
 CC protein encoded by the insecticidal secreted protein (isp1a or isp2a) DNA
 CC deposited at the BCM-LMBP under accession number LMBP 4009. Also
 CC included are a DNA sequence encoding the protein cited above, a chimaeric
 CC gene comprising the DNA and a plant-expressible promoter region (e.g. the
 CC long and short forms of the corn zrp2 gene promoter), a plant cell
 CC comprising the chimaeric gene used to produce the protein cited above, a
 CC plant or a seed comprising the chimeric gene of integrated in its cells,
 CC a microorganism transformed to contain the DNA encoding ISP1A or ISP2A, a
 CC process for controlling insects comprising expressing the protein cited
 CC above in cells of a plant, a process for rendering a plant resistant to
 CC coleopteran insects (comprising transforming plant cells with a first
 CC chimaeric gene and a second chimeric gene, and regenerating transformed
 CC plants from the cells which are resistant to insects, where the first and
 CC second chimaeric genes are ISP1A and ISP2A, and a process for controlling
 CC coleopteran insect pests (comprising planting, sowing or growing in a
 CC field plants transformed with a first chimaeric gene and a second
 CC chimaeric gene, where the first and second chimaeric genes are ISP1A and
 CC ISP2A). The insecticidal secreted proteins and DNA sequences are useful
 CC in making insecticidal compositions or transgenic plants to protect
 CC plants (especially crops like corn) from damage caused by insects,
 CC particularly coleopteran insects (e.g. Diabrotica sp., Leptinotarsa sp.
 CC and Anthonomus sp.). The proteins are useful for preventing or minimizing
 CC insect damage to plants in a field. The present sequence represents a
 CC truncated ISP1A protein where the signal peptide has been replaced by Met
 CC -Ala
 XX
 XX Sequence 836 AA;
 SQ
 Query Match 6.3%; Score 96.5; DB 6; Length 836;
 Best Local Similarity 21.0%; Pred. No. 11;
 Matches 64; Conservative 48; Mismatches 140; Indels 53; Gaps 14;
 QY 8 LAAMVALAAPVQSVVFAQT---NNSESAPILRWASAEKHNHSHLWVNRADIM-- 62
 DB 267 VAAFPVNVVLEKVLKSKNEDLSHVSQSTNWSY--TNTGVN-----VNAGWSGLGP 319
 QY 63 SRNTIVNPNETALNEWADLENGIYSADYENPY-----YDN-STYASHFYDDPT--- 112
 DB 320 SFGVSVNYQHSETVANEGSATNDGTHNGASAYLNANRYNNVGTGAIYETKTTSPF 379
 QY 113 --GTTTIPAKHAKETGAKYFNLAGQYQNMQQAFFYGLSHLYLGDVNPMPHAANPT 170
 DB 380 LGDTT-IGTIK-AKENTALTILPDQSYPEKGN-----GIAINTMDDFNSRPPLNKE 431
 QY 171 NLSYPMGFHSKYENFVDTTKNNYIVSDSG-----YNNWKGANPEWIEGAIV----- 218
 DB 432 QLNTYLSNKKPILLETDOVEGKYAIKDTNGNTIAGDMNGITDEISAKTASIIVDNGNQ 491

QY 219 ----AAKODYPGVVNDTTKDMFVKAVSQBYADKWRAEVTPVTGKRLMEARVTVAGIHL 274
 DB 492 SEKRVAAKDTN-PEDKTPNLSVKALKLAYPD-----EIEKDGLLFYNDQPIFEASVQS 546
 QY 275 WFDTY 279
 DB 547 YVDEY 551

RESULT 9 AAG68285

ID AAG68285 standard; protein; 871 AA.

XX AC AAG68285;

XX DT 19-FEB-2002 (first entry)

XX DE Brevibacillus laterosporus ISP1A protein SEQ ID NO:2.

XX KW Brevibacillus laterosporus; insecticidal secreted protein; ISP1A; ISP2A;
 XX ISP1A-1; ISP2A-1; toxin; bacterial; plant protection.

XX OS Brevibacillus laterosporus.

XX FH Key Location/Qualifiers

FT Peptide 1..37

FT Protein /label= signal

FT 38..871

FT /label= ISP1A

XX WO200187931-A2.

XX PD 22-NOV-2001.

XX PF 17-MAY-2001; 2001WO-EP005702.

XX PR 18-MAY-2000; 2000US-00573872.

XX XX (AVET) AVENTIS CROPS SCIENCE NV.

XX PI Boets A, Arnaut G, Damme N, Van Rie J;

XX DR WPI; 2002-062371/08.

XX DR N-PSDB; ABA04056.

XX XX New bacterial insecticidal proteins, useful for making insecticidal
 PT composition to protect plants from damage by insects especially
 PT coleopteran insects.

XX PS Claim 4; Page 61-65; 86pp; English.

XX CC The present sequence represents a Brevibacillus laterosporus insecticidal
 CC secreted protein designated ISP1A. ISP1A, and ISP2A, are insecticidal to
 CC Diabrotica virgifera larvae when ingested by larvae in combination with
 CC the ISP2A amino acid sequence 51 to 457, or ISP1A amino acid sequence 38
 CC to 871, respectively. ISP1A and ISP2A are useful for controlling insects.
 CC The polynucleotide sequences encoding ISP1A and ISP2A can be used in the
 CC production of transgenic plants resistant to coleopteran insects

XX SQ Sequence 871 AA;

Query Match 6.3%; Score 96.5; DB 5; Length 871;

Best Local Similarity 21.0%; Pred. No. 12;

Matches 64; Conservative 48; Mismatches 140; Indels 53; Gaps 14;

QY 8 LAAMVALAAPVQSVVFAQT---NNSESAPILRWASAEKHNHSHLWVNRADIM-- 62

DB 302 VAAFPVNVVLEKVLKSKNEDLSHVSQSTNWSY--TNTGVN-----VNAGWSGLGP 354

QY 63 SRNTIVNPNETALNEWADLENGIYSADYENPY-----YDN-STYASHFYDDPT--- 112

DB 355 SFGVSVNYQHSETVANEGSATNDGTHNGASAYLNANRYNNVGTGAIYETKTTSPF 414

QY 113 --GTTTIPFAKHAKETGAKYFNLAGQAYONQDQAFYFGLSLHLYGDNVQPMHAAAPT 170
DB 415 LDGTT-IGTIK-AKENTTALTILPDQSYPERGKN-----GIAINTMDDFNSRPIPLNKE 466
QY 171 NLSYPMGFHSKYENFVDTIKNNYIVSDSG-----YNNWKGANPEDWIEGAAV----- 218
DB 467 QLNTYLSNKKPILLETQVEGKYAIKDTNGNITAGDWNIGITDEISAKTASIIVDNGNQ 526
QY 219 ----AAKQDYPGVNDTTKDMFVKAASQYADKWRAEVTPVTGKRLMEAGQRTVAGYIHL 274
DB 527 SEKRVAAKDYTN-PEDKTPNLSVKEALKLAYPD-----EIEEKDGLLFYNDQPIFEASVQS 581
QY 275 WFDY 279
DB 582 YVDEY 586

RESULT 10
ABU08981
ID ABU08981 standard; protein; 871 AA.
XX AC ABU08981;
XX AC
XX DT 16-JUN-2003 (first entry)
XX DE B. laterosporus insecticidal secreted protein, ISPIA.
XX KW Insecticidal secreted protein; toxin; ISPIA; insecticide; Diabrotica sp;
XX KW Leptinotarsa sp; Anthonomus sp; corn; coleopteran insect; transgenic.
XX OS Brevibacillus laterosporus.
XX FH Key Location/Qualifiers
XX FT Peptide 1..37
XX FT /label= Signal_peptide
XX FT Protein 38..871
XX FT /label= Mature_ISPIA
XX FT /note= "This protein is specifically claimed in claim 2"
XX FT Protein 38..768
XX FT /label= Active_fragment
XX FT /note= "This protein is specifically claimed in claim 3"
XX FN US2002199215-A1.
XX PD 26-DEC-2002.
XX PF 17-MAY-2001; 2001US-00858525.
XX PR 17-MAY-2001; 2001US-00858525.
XX PR (BOET/) BOETS A.
XX PA (ARNA/) ARNAUT G.
XX PA (RIEJ/) RIE J V.
XX PA (DAMW/) DAME N.
XX XX
XX PI Boets A, Arnaut G, Rie JV, Damme N;
XX XX
XX DR WPI: 2003-361872/34.
XX DR N-PSDB; ABX94348.
XX XX
XX FT New insecticidal secreted proteins and DNA sequences, useful in making
XX FT insecticidal compositions or transgenic plants to protect plants from
XX FT damage caused by insects, particularly coleopteran insects.
XX XX
XX PS Claim 1; Page 18-20; 38pp; English.
XX CC
XX CC The invention relates to a new insecticidal secreted protein comprising:
XX CC (a) amino acid sequence of the smallest active toxin of the protein
XX CC comprising a sequence appearing as ABU08981 (ISPIA) or ABU08982 (ISP2A);
XX CC or (b) the amino acid sequence of the protease-digestion fragment of the
XX CC protein encoded by the insecticidal secreted protein (ispiA or isp2A DNA
XX CC deposited at the BCCM-LMBP under accession number LMBP 4009. Also

CC included are a DNA sequence encoding the protein cited above, a chimaeric
CC gene comprising the DNA and a plant-expressible promoter region (e.g. the
CC long and short forms of the corn zip2 gene promoter), a plant cell
CC comprising the chimaeric gene used to produce the protein cited above, a
CC plant or a seed comprising the chimeric gene of integrated in its cells,
CC a microorganism transformed to contain the DNA encoding ISPIA or ISP2A, a
CC process for controlling insects comprising expressing the protein cited
CC above in cells of a plant, a process for rendering a plant resistant to
CC coleopteran insects (comprising transforming plant cells with a first
CC chimaeric gene and a second chimeric gene, and regenerating transformed
CC plants from the cells which are resistant to insects, where the first and
CC second chimaeric genes are ISPIA and ISP2A, and a process for controlling
CC coleopteran insect pests (comprising planting, sowing or growing in a
CC field plants transformed with a first chimaeric gene and a second
CC chimaeric gene, where the first and second chimaeric genes are ISPIA and
CC ISP2A). The insecticidal secreted proteins and DNA sequences are useful
CC in making insecticidal compositions or transgenic plants to protect
CC plants (especially crops like corn) from damage caused by insects,
CC particularly coleopteran insects (e.g. Diabrotica sp., Leptinotarsa sp.
CC and Anthonomus sp.). The proteins are useful for preventing or minimising
CC insect damage to plants in a field. The present sequence represents ISPIA
CC protein
XX

SQ Sequence 871 AA;

Query Match 6.3%; Score 96.5; DB 6; Length 871;
Best Local Similarity 21.0%; Pred. No. 12;
Matches 64; Conservative 48; Mismatches 140; Indels 53; Gaps 14;

QY 8 LAAMVALAAPQSVVFAQT---NNSESPAPILRWSAEDKHNEGINSHLWIVNRAIDIM-- 62
DB 302 VAAPSVNVSLKVLKSNEDLSHVSSESQSTNWSY--TNTGCVN-----VNAGWSGLGP 354
QY 63 SRNTTIVNPNETALINEWRADLENGIYSADYENPY-----YDN-STYASHFYDPTDPT--- 112
DB 355 SFGVSVNYQHSFETVANEGSATNDGTHNGAESAYLANAVYNNVGTGAIYETKTTSTFI 414
QY 113 --GTTTIPFAKHAKETGAKYFNLAGQAYONQDQAFYFGLSLHLYGDNVQPMHAAAPT 170
DB 415 LDGTT-IGTIK-AKENTTALTILPDQSYPERGKN-----GIAINTMDDFNSRPIPLNKE 466
QY 171 NLSYPMGFHSKYENFVDTIKNNYIVSDSG-----YNNWKGANPEDWIEGAAV----- 218
DB 467 QLNTYLSNKKPILLETQVEGKYAIKDTNGNITAGDWNIGITDEISAKTASIIVDNGNQ 526
QY 219 ----AAKQDYPGVNDTTKDMFVKAASQYADKWRAEVTPVTGKRLMEAGQRTVAGYIHL 274
DB 527 SEKRVAAKDYTN-PEDKTPNLSVKEALKLAYPD-----EIEEKDGLLFYNDQPIFEASVQS 581
QY 275 WFDY 279
DB 582 YVDEY 586

RESULT 11

ABR52701
ID ABR52701 standard; protein; 1283 AA.
XX AC ABR52701;
XX AC

XX DT 20-JUN-2003 (first entry)

XX DE Protein sequence #SEQ ID 267.

XX KW Multiprotein complex; eukaryote; drug target; diagnosis.

XX OS Saccharomyces cerevisiae.

XX FN EP1258494-A1.

XX PD 20-NOV-2002.

XX PF 20-DEC-2001; 2001EP-00130253.

XX 15-MAY-2001; 2001EP-00111774.
XX (CELL-) CELLZOME AG.
XX Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;
PI Marzioch M, Schultz JD, Superti-Furga GD;
XX WPI; 2003-250078/25.
XX N-PSDB; ACC60743.
XX New isolated protein complexes useful for diagnosing a disease or
XX disorder, or as a target for an active agent of a pharmaceutical,
XX preferably a drug target in the treatment or prevention of disease or
XX disorder.
XX Disclosure; SEQ ID NO 267; 17pp + Sequence Listing; English.
XX
XX The invention relates to multiprotein complexes from eukaryotes. Proteins
XX of the invention and DNA sequences encoding them are given in records
XX ABR52568-ABR53903 and ACC6010-ACC61944 respectively. The complexes are
XX obtainable by using a protein as a bait and isolating the set of proteins
XX which is attached thereto from cells. Such protein complexes may comprise
XX up to 30 distinct proteins. Protein complexes of the invention are useful
XX for diagnosing a disease or disorder, or as a target for an active agent
XX of a pharmaceutical, preferably a drug target in the treatment or
XX prevention of a disease or disorder. Note: The sequence data for this
XX patent is not represented in the printed specification, but is based on
XX sequence information supplied by the European Patent Office. The complete
XX document is available on CD-ROM
XX
XX Sequence 1283 AA;
Query Match 6.3%; Score 95.5; DB 6; Length 1283;
Best Local Similarity 25.9%; Pred. No. 26;
Matches 72; Conservative 31; Mismatches 104; Indels 71; Gaps 17;
QY 1 MKKVLALAAVLAAPVQSVVFAQTNNSESPAPILRWSAEDKHNEGINSH----- 51
DB 838 VKEKTEKIASSVIGETKIIVAVTVQKKEEY---LKEGSLYGEDGKRLSMDKDRP 894
QY 52 ---LWIVNRAI--DIMSRNTTIVNPNE-TALLNEWRADLENGIYSADYENFYDNTSVAS 105
DB 895 KISLMAVLKSMVKDMTRTLPTFTNEPTSLQKRVADLE-----YSELDDQAA 943
QY 106 HFVDPDTGTYYI-----PFAKAKETGAKYFN-LAGQA--YQNDMQQAFFYGLSLHY 156
DB 944 TFEDSTLRLTYAAFTASSYASTTKRV-AKPNPLIGETFEYSRDPKQYRFFTEQVSHH- 1001
QY 157 LGDVNCPMAAFTNLSPMGPHSKYENFVDTIKNNYIVSDGN-GYWN----- 203
DB 1002 -----PPISATWTE-SPWDFWG--ESFVDTKFNGRSFNVKHLGLWHIKLRPNDEKEE 1052
QY 204 ---WKGANPDMTEGAARAAKQ-DYPG---VVDNDDTKD 234
DB 1053 LITWK--KPNNTVIGILIGNPQVDNHRGEVNVVNHHTGD 1088
RESULT 12
AAG52327
ID AAG52327 standard; protein; 192 AA.
XX AAG52327;
AC AAG52327;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 66504.
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.

XX PN EP1033405-A2.
XX XX
XX PD 06-SEP-2000.
XX XX
XX 25-FEB-2000; 2000EP-00301439.
XX PF 99US-0121825P.
XX PR 99US-0123180P.
XX PR 05-MAR-1999; 99US-0123180P.
XX PR 09-MAR-1999; 99US-0123180P.
XX PR 23-MAR-1999; 99US-0125788P.
XX PR 25-MAR-1999; 99US-0126264P.
XX PR 29-MAR-1999; 99US-0126785P.
XX PR 01-APR-1999; 99US-0127462P.
XX PR 06-APR-1999; 99US-0128234P.
XX PR 08-APR-1999; 99US-0128714P.
XX PR 16-APR-1999; 99US-0129645P.
XX PR 19-APR-1999; 99US-0130077P.
XX PR 21-APR-1999; 99US-0130449P.
XX PR 23-APR-1999; 99US-0130510P.
XX PR 28-APR-1999; 99US-0130891P.
XX PR 30-APR-1999; 99US-0131449P.
XX PR 30-APR-1999; 99US-0132048P.
XX PR 04-MAY-1999; 99US-0132407P.
XX PR 05-MAY-1999; 99US-0132484P.
XX PR 06-MAY-1999; 99US-0132485P.
XX PR 06-MAY-1999; 99US-0132486P.
XX PR 07-MAY-1999; 99US-0132487P.
XX PR 11-MAY-1999; 99US-0132863P.
XX PR 14-MAY-1999; 99US-0134256P.
XX PR 14-MAY-1999; 99US-0134218P.
XX PR 14-MAY-1999; 99US-0134219P.
XX PR 14-MAY-1999; 99US-0134221P.
XX PR 14-MAY-1999; 99US-0134370P.
XX PR 18-MAY-1999; 99US-0134768P.
XX PR 19-MAY-1999; 99US-0134941P.
XX PR 20-MAY-1999; 99US-0135124P.
XX PR 21-MAY-1999; 99US-0135353P.
XX PR 24-MAY-1999; 99US-0135629P.
XX PR 25-MAY-1999; 99US-0136021P.
XX PR 27-MAY-1999; 99US-0136392P.
XX PR 28-MAY-1999; 99US-0136782P.
XX PR 01-JUN-1999; 99US-0137222P.
XX PR 03-JUN-1999; 99US-0137528P.
XX PR 04-JUN-1999; 99US-0137502P.
XX PR 07-JUN-1999; 99US-0137724P.
XX PR 08-JUN-1999; 99US-0138094P.
XX PR 10-JUN-1999; 99US-0138540P.
XX PR 10-JUN-1999; 99US-0138847P.
XX PR 14-JUN-1999; 99US-0139119P.
XX PR 16-JUN-1999; 99US-0139452P.
XX PR 16-JUN-1999; 99US-0139453P.
XX PR 17-JUN-1999; 99US-0139492P.
XX PR 18-JUN-1999; 99US-0139454P.
XX PR 18-JUN-1999; 99US-0139455P.
XX PR 18-JUN-1999; 99US-0139456P.
XX PR 18-JUN-1999; 99US-0139457P.
XX PR 18-JUN-1999; 99US-0139458P.
XX PR 18-JUN-1999; 99US-0139459P.
XX PR 18-JUN-1999; 99US-0139460P.
XX PR 18-JUN-1999; 99US-0139461P.
XX PR 18-JUN-1999; 99US-0139462P.
XX PR 18-JUN-1999; 99US-0139463P.
XX PR 18-JUN-1999; 99US-0139750P.
XX PR 18-JUN-1999; 99US-0139763P.
XX PR 21-JUN-1999; 99US-0139817P.
XX PR 22-JUN-1999; 99US-0139899P.
XX PR 23-JUN-1999; 99US-0140353P.
XX PR 23-JUN-1999; 99US-0140354P.
XX PR 24-JUN-1999; 99US-0140695P.
XX PR 28-JUN-1999; 99US-0140823P.
XX PR 29-JUN-1999; 99US-0140991P.
XX PR 30-JUN-1999; 99US-0141287P.

PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144086P.
PR 16-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145113P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145911P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149923P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151068P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.

PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 6.2%; Score 94.5; DB 3; Length 192;

Best Local Similarity 21.3%; Pred. No. 2;

Matches 35; Conservative 29; Mismatches 49; Indels 51; Gaps 9;

QY 135 GQAYNQDMQQAFFYGLSLHVLGDNOPMHAANFTN-----LSYPMGFH---- 179

Db 20 GTSRRRYNMTALLFLS---HFMGDIHQPMH-VGFTSDEGGNTIDLRWYKHKSNLHHWVD 75

QY 180 -----SKYENFVDTKNNYIVSDSNGYW-----NWKGAN-----PEDWI-ECAAVA 219

Db 76 REIILTALKENYDNLDLLOEDLEKNTITGLHDDLSWTECNDLIACPHKYASISIKLA 135

QY 220 AKQDYPGVNDTTKDMFVKAASOEYADKWAETVPTVGKRLME 263

Db 136 CKWGYKGVKSGET-----LSBEYFN-----TRLPIVMKRIQV 167

RESULT 13

AAG05009

ID AAG05009 standard; protein; 192 AA.

XX AAG05009;

AC AAG05009;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 1251.

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX

OS	Arabidopsis thaliana.		
XX	PN	EP1033405-A2.	PR
XX	XX		PR
XX	PD	06-SEP-2000.	PR
XX	PF	25-FEB-2000; 2000EP-00301439.	PR
XX	XX		PR
XX	XX	25-FEB-1999;	PR
XX	XX	99US-0121825P.	99US-0141287P.
XX	XX	05-MAR-1999;	99US-0141842P.
XX	XX	99US-0123180P.	99US-0142154P.
XX	XX	09-MAR-1999;	99US-0142055P.
XX	XX	99US-0123548P.	99US-0142390P.
XX	XX	23-MAR-1999;	99US-0142803P.
XX	XX	99US-0125788P.	99US-0142920P.
XX	XX	25-MAR-1999;	99US-0142977P.
XX	XX	99US-0126264P.	99US-0143542P.
XX	XX	29-MAR-1999;	99US-0143624P.
XX	XX	99US-0126785P.	99US-0144005P.
XX	XX	01-APR-1999;	99US-0144086P.
XX	XX	99US-0127462P.	99US-0144325P.
XX	XX	06-APR-1999;	99US-0144331P.
XX	XX	99US-0128234P.	99US-0144332P.
XX	XX	08-APR-1999;	99US-0144333P.
XX	XX	99US-0128714P.	99US-0144334P.
XX	XX	16-APR-1999;	99US-0144335P.
XX	XX	99US-0129845P.	99US-0144352P.
XX	XX	19-APR-1999;	99US-0144632P.
XX	XX	99US-0130077P.	99US-0144884P.
XX	XX	21-APR-1999;	99US-0144814P.
XX	XX	99US-0130444P.	99US-0145086P.
XX	XX	23-APR-1999;	99US-0145088P.
XX	XX	99US-0130510P.	99US-0145085P.
XX	XX	28-APR-1999;	99US-0145087P.
XX	XX	99US-0130891P.	99US-0145089P.
XX	XX	30-APR-1999;	99US-0145192P.
XX	XX	99US-0131149P.	99US-0145145P.
XX	XX	03-MAY-1999;	99US-0145218P.
XX	XX	99US-0132049P.	99US-0145224P.
XX	XX	30-APR-1999;	99US-0145913P.
XX	XX	99US-0132407P.	99US-0145918P.
XX	XX	04-MAY-1999;	99US-0145919P.
XX	XX	99US-0132484P.	99US-0145918P.
XX	XX	05-MAY-1999;	99US-0145919P.
XX	XX	99US-0132486P.	99US-0146386P.
XX	XX	06-MAY-1999;	99US-0146388P.
XX	XX	99US-0132487P.	99US-0146389P.
XX	XX	07-MAY-1999;	99US-0147038P.
XX	XX	99US-0132863P.	99US-0147204P.
XX	XX	11-MAY-1999;	99US-0147302P.
XX	XX	99US-0134256P.	99US-0147192P.
XX	XX	14-MAY-1999;	99US-0147260P.
XX	XX	99US-0134218P.	99US-0147303P.
XX	XX	14-MAY-1999;	99US-0147416P.
XX	XX	99US-0134370P.	99US-0147493P.
XX	XX	18-MAY-1999;	99US-0147935P.
XX	XX	99US-0134768P.	99US-0148171P.
XX	XX	19-MAY-1999;	99US-0148319P.
XX	XX	99US-0134941P.	99US-0148341P.
XX	XX	20-MAY-1999;	99US-0148565P.
XX	XX	99US-0135124P.	99US-0148684P.
XX	XX	21-MAY-1999;	99US-0149368P.
XX	XX	99US-0135353P.	99US-0149175P.
XX	XX	24-MAY-1999;	99US-0149262P.
XX	XX	99US-0135629P.	99US-0149222P.
XX	XX	25-MAY-1999;	99US-0149723P.
XX	XX	99US-0136021P.	99US-0149729P.
XX	XX	27-MAY-1999;	99US-0149902P.
XX	XX	99US-0136392P.	99US-0149930P.
XX	XX	28-MAY-1999;	99US-0150566P.
XX	XX	99US-0136782P.	99US-0150884P.
XX	XX	01-JUN-1999;	99US-0151065P.
XX	XX	99US-0137222P.	99US-0151066P.
XX	XX	04-JUN-1999;	99US-0151080P.
XX	XX	99US-0137502P.	99US-0151303P.
XX	XX	07-JUN-1999;	99US-0151438P.
XX	XX	99US-0137724P.	99US-0151930P.
XX	XX	08-JUN-1999;	99US-0152363P.
XX	XX	99US-0138094P.	99US-0153070P.
XX	XX	10-JUN-1999;	99US-0153758P.
XX	XX	99US-0138540P.	
XX	XX	14-JUN-1999;	
XX	XX	99US-0139119P.	
XX	XX	16-JUN-1999;	
XX	XX	99US-0139452P.	
XX	XX	16-JUN-1999;	
XX	XX	99US-0139453P.	
XX	XX	17-JUN-1999;	
XX	XX	99US-0139492P.	
XX	XX	18-JUN-1999;	
XX	XX	99US-0139454P.	
XX	XX	18-JUN-1999;	
XX	XX	99US-0139455P.	
XX	XX	18-JUN-1999;	
XX	XX	99US-0139456P.	
XX	XX	18-JUN-1999;	
XX	XX	99US-0139457P.	
XX	XX	18-JUN-1999;	
XX	XX	99US-0139458P.	
XX	XX	18-JUN-1999;	
XX	XX	99US-0139459P.	
XX	XX	18-JUN-1999;	
XX	XX	99US-0139460P.	
XX	XX	18-JUN-1999;	
XX	XX	99US-0139461P.	
XX	XX	18-JUN-1999;	
XX	XX	99US-0139462P.	
XX	XX	18-JUN-1999;	
XX	XX	99US-0139463P.	
XX	XX	18-JUN-1999;	
XX	XX	99US-0139750P.	
XX	XX	18-JUN-1999;	
XX	XX	99US-0139763P.	
XX	XX	21-JUN-1999;	
XX	XX	99US-0139817P.	
XX	XX	22-JUN-1999;	
XX	XX	99US-0139899P.	
XX	XX	23-JUN-1999;	
XX	XX	99US-0140353P.	
XX	XX	23-JUN-1999;	
XX	XX	99US-0140354P.	
XX	XX	24-JUN-1999;	
XX	XX	99US-0140655P.	
XX	XX	28-JUN-1999;	
XX	XX	99US-0140823P.	
XX	XX	29-JUN-1999;	
XX	XX	99US-0140991P.	


```
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 19-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147418P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149923P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.

PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158023P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159323P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 6.2%; Score 94.5; DB 3; Length 305;
Best Local Similarity 21.3%; Pred. No. 3.9;
Matches 35; Conservative 29; Mismatches 49; Indels 51; Gaps 9;

Qy 135 GOAYQNQDMQQAFFYLGSLHLYLGDVNPQMHAANFTN-----LSYPMGFH----- 179
Db 133 GTSRRYNNMTEALLFLS---HFMDIHQPMH-VGFTSDEGGNTIDLWYKXKSNLHHWD 188
Qy 180 -----SKYENFVDTIKNNYIVSDSNGW-----NWKGAN-----PEDMI-EGAAVA 219
Db 189 REIILTALKENYDKNLDLQEDLEKNTINGLHMDLSSWTECNLDLIACPHKYASESIKLA 248
Qy 220 AKQDYPGVNDTTKDMFVKAASVSOEYADKRAEVTPTVTKRLME 263
Db 249 CKWGYKGVKSGET-----LSEYFN-----TFLPIVMKRIVQ 280

RESULT 15
AAG52326
ID AAG52326 standard; protein; 305 AA.
XX
AC AAG52326;
XX
DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 66503.
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
KW
```

KW		termination sequence.	
XX	OS	<i>Arabidopsis thaliana</i> .	
XX	PN	EP1033405-A2.	
XX	PD	06-SEP-2000.	
XX	PF	25-FEB-2000; 2000EP-0301439.	
XX	PR	25-FEB-1999;	99US-0128125P.
PR	PR	05-MAR-1999;	99US-0123180P.
PR	PR	09-MAR-1999;	99US-0123548P.
PR	PR	23-MAR-1999;	99US-0125788P.
PR	PR	23-MAR-1999;	99US-0126264P.
PR	PR	29-MAR-1999;	99US-0126785P.
PR	PR	01-APR-1999;	99US-0127462P.
PR	PR	06-APR-1999;	99US-0128234P.
PR	PR	08-APR-1999;	99US-0128714P.
PR	PR	16-APR-1999;	99US-0129845P.
PR	PR	19-APR-1999;	99US-0130077P.
PR	PR	21-APR-1999;	99US-0130449P.
PR	PR	23-APR-1999;	99US-0130510P.
PR	PR	28-APR-1999;	99US-0130891P.
PR	PR	30-APR-1999;	99US-0131449P.
PR	PR	30-APR-1999;	99US-0132048P.
PR	PR	04-MAY-1999;	99US-0132407P.
PR	PR	05-MAY-1999;	99US-0132484P.
PR	PR	06-MAY-1999;	99US-0132486P.
PR	PR	06-MAY-1999;	99US-0132487P.
PR	PR	07-MAY-1999;	99US-0132863P.
PR	PR	11-MAY-1999;	99US-0134256P.
PR	PR	14-MAY-1999;	99US-0134218P.
PR	PR	14-MAY-1999;	99US-0134219P.
PR	PR	14-MAY-1999;	99US-0134221P.
PR	PR	14-MAY-1999;	99US-0134370P.
PR	PR	18-MAY-1999;	99US-0134768P.
PR	PR	19-MAY-1999;	99US-0134961P.
PR	PR	20-MAY-1999;	99US-0135124P.
PR	PR	21-MAY-1999;	99US-0135353P.
PR	PR	24-MAY-1999;	99US-0135829P.
PR	PR	25-MAY-1999;	99US-0136021P.
PR	PR	27-MAY-1999;	99US-0136392P.
PR	PR	28-MAY-1999;	99US-0136782P.
PR	PR	01-JUN-1999;	99US-0137222P.
PR	PR	03-JUN-1999;	99US-0137528P.
PR	PR	04-JUN-1999;	99US-0137502P.
PR	PR	07-JUN-1999;	99US-0137724P.
PR	PR	08-JUN-1999;	99US-0138094P.
PR	PR	10-JUN-1999;	99US-0138540P.
PR	PR	10-JUN-1999;	99US-0138847P.
PR	PR	14-JUN-1999;	99US-0139119P.
PR	PR	16-JUN-1999;	99US-0139452P.
PR	PR	16-JUN-1999;	99US-0139453P.
PR	PR	17-JUN-1999;	99US-0139492P.
PR	PR	18-JUN-1999;	99US-0139454P.
PR	PR	18-JUN-1999;	99US-0139455P.
PR	PR	18-JUN-1999;	99US-0139456P.
PR	PR	18-JUN-1999;	99US-0139457P.
PR	PR	18-JUN-1999;	99US-0139458P.
PR	PR	18-JUN-1999;	99US-0139459P.
PR	PR	18-JUN-1999;	99US-0139460P.
PR	PR	18-JUN-1999;	99US-0139461P.
PR	PR	18-JUN-1999;	99US-0139462P.
PR	PR	18-JUN-1999;	99US-0139463P.
PR	PR	18-JUN-1999;	99US-0139750P.
PR	PR	18-JUN-1999;	99US-0139763P.
PR	PR	21-JUN-1999;	99US-0139817P.
PR	PR	23-JUN-1999;	99US-0139899P.
PR	PR	23-JUN-1999;	99US-0140353P.
PR	PR	23-JUN-1999;	99US-0140354P.
PR	PR	24-JUN-1999;	99US-0140695P.
99US-0140823P.	PR	28-JUN-1999;	
99US-0140991P.	PR	29-JUN-1999;	
99US-01411287P.	PR	30-JUN-1999;	
99US-01411842P.	PR	01-JUL-1999;	
99US-0142154P.	PR	01-JUL-1999;	
99US-0142055P.	PR	02-JUL-1999;	
99US-0142390P.	PR	06-JUL-1999;	
99US-0142803P.	PR	08-JUL-1999;	
99US-0142920P.	PR	09-JUL-1999;	
99US-0142977P.	PR	12-JUL-1999;	
99US-0143542P.	PR	13-JUL-1999;	
99US-0143624P.	PR	14-JUL-1999;	
99US-0144005P.	PR	15-JUL-1999;	
99US-0144085P.	PR	16-JUL-1999;	
99US-0144086P.	PR	16-JUL-1999;	
99US-0144325P.	PR	19-JUL-1999;	
99US-0144331P.	PR	19-JUL-1999;	
99US-0144332P.	PR	19-JUL-1999;	
99US-0144333P.	PR	19-JUL-1999;	
99US-0144334P.	PR	19-JUL-1999;	

```
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 6.2%; Score 94.5; DB 3; Length 305;
Best Local Similarity 21.3%; Pred. No. 3.9;
Matches 35; Conservative 29; Mismatches 49; Indels 51; Gaps 9;

QY 135 GOAYQNDMQCAFFYGLSLHYLDVNPQMHAAFTN-----LSYPMGFH--- 179
Db 133 GTSRRYNTALLFLS---HFMGDIHQPMH-VGFTSDEGGNTIDLRWYKXSNLHHYWD 188
QY 180 -----SKYENFVDTIKNYIYVDSNGYW-----NWKGAN-----PEDWI-EGAAVA 219
Db 189 REIILTALKENYDKNLDLLQEDLEKNIITNGLWHDDLSSWTECNDLIACPHKYASESIKLA 248
QY 220 AKQDYPGVNDTTKDFVKAASQYADKWRAEVTPVTGKRLME 263
Db 249 CKWGYKGVKSGET-----LSEYFN-----TRLPIVMKRIVQ 280
```

Search completed: July 27, 2004, 14:46:07
Job time : 56 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2004, 14:45:12 ; Search time 18 Seconds
(without alignments)
808.807 Million cell updates/sec

Title: PCT-US03-12556-2
Perfect score: 1522
Sequence: 1 MKKKVLLAMVALAPVQS.....EAQRTAGYIHLNFDYVNR 282

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA*
1: /cgm2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgm2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgm2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgm2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgm2_6/ptodata/2/iaa/PCUS_COMB.pep:*
6: /cgm2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	220	14.5	370	2	US-08-341-538A-2
2	220	14.5	370	2	US-08-725-518-2
3	96.5	6.3	276	4	US-09-510-322A-11
4	95	6.2	274	3	US-08-491-954-4
5	94	6.2	399	2	US-08-282-197C-54
6	94	6.2	685	4	US-09-252-991A-32033
7	94	6.2	837	4	US-09-390-234-14
8	94	6.2	837	4	US-09-603-311-14
9	93	6.1	275	4	US-09-510-322A-10
10	91.5	6.0	372	4	US-09-092-315-13
11	91.5	6.0	464	4	US-09-092-315-1
12	91.5	6.0	464	4	US-09-733-524A-1
13	91.5	6.0	478	4	US-09-092-315-7
14	91.5	6.0	478	4	US-09-733-524A-7
15	91.5	6.0	503	4	US-09-252-991A-17709
16	90	5.9	575	4	US-09-134-001C-3906
17	90	5.9	727	4	US-09-134-001C-4067
18	89.5	5.9	636	3	US-08-737-752A-2
19	89	5.8	218	4	US-09-615-192A-269
20	89	5.8	474	4	US-09-489-039A-13140
21	89	5.8	987	4	US-09-543-681A-7785
22	88.5	5.8	378	4	US-09-489-039A-13960
23	88.5	5.8	423	4	US-09-328-352-5224
24	87.5	5.7	493	4	US-09-543-681A-5068
25	87.5	5.7	515	4	US-09-107-532A-5603
26	87	5.7	325	4	US-09-543-681A-7453
27	86.5	5.7	291	4	US-09-495-406-5

28 86.5 5.7 464 3 US-09-426-072-2 Sequence 2, Appli
29 86 5.7 296 4 US-09-510-322A-2 Sequence 2, Appli
30 85.5 5.6 291 4 US-09-495-406-9 Sequence 9, Appli
31 85 5.6 890 2 US-08-483-101-14 Sequence 14, Appl
32 84.5 5.6 713 4 US-09-489-039A-10668 Sequence 10668, A
33 83.5 5.5 268 4 US-09-134-000C-3519 Sequence 3519, Ap
34 83.5 5.5 418 4 US-09-958-548-1 Sequence 1, Appli
35 83.5 5.5 731 4 US-09-107-532A-6999 Sequence 6999, Ap
36 83.5 5.5 901 4 US-09-134-001C-5389 Sequence 5389, Ap
37 83 5.5 511 1 US-08-480-604A-20 Sequence 20, Appl
38 83 5.5 511 2 US-08-405-496A-20 Sequence 20, Appl
39 83 5.5 511 3 US-08-915-136-20 Sequence 20, Appl
40 83 5.5 511 4 US-08-957-310-20 Sequence 20, Appl
41 83 5.5 511 4 US-10-011-366-20 Sequence 20, Appl
42 83 5.5 511 4 US-09-084-517-20 Sequence 20, Appl
43 83 5.5 608 1 US-08-480-604A-21 Sequence 21, Appl
44 83 5.5 608 2 US-08-405-496A-21 Sequence 21, Appl
45 83 5.5 608 3 US-08-915-136-21 Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-08-341-538A-2
; Sequence 2, Application US/08341538A
; Patent No. 5817317
; GENERAL INFORMATION:
; APPLICANT: TITBALL DR, RICHARD W
; APPLICANT: WILLIAMSON DR, ETHEL D
; NUMBER OF INVENTION: CLOSTRIDIUM PERFRINGENS VACCINES
; TITLE OF INVENTION: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE
; STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: UNITED STATES
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/341,538A
; FILING DATE: 28-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9210717.6
; FILING DATE: 20-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9215655.3
; FILING DATE: 23-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO GB93/01039
; FILING DATE: 20-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: CRAWFORD, ARTHUR R
; REGISTRATION NUMBER: 25327
; REFERENCE/DOCKET NUMBER: 1498-58
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 370 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-341-538A-2

Query Match 14.5%; Score 220; DB 2; Length 370;
Best Local Similarity 26.5%; Pred. No. 6.8e-14;
Matches 68; Conservative 40; Mismatches 95; Indels 54; Gaps 10;

```

QY 46 EGINSHLWVNRADIMSRNTTIVNPNET-----ALLNWRADLENGIYSADYENPYDNS 101
D 6 DGTGTHAMIVTQGVSIENDLSKNEPESVRKNLEILKENMHELQLGSTPYDYNKAYD-- 63
QY 102 TYASHFYDDPTGTTTYP-----FAKHAKETG-----AKYFNLAGOAYQNQDMQOAFYYLGL 152
D 64 LYQDFHWDPTDNNFNSKDNSWYLAYSIDPTGESQIRKFSALARYEWQRGNYKQATFYLGE 123
QY 153 SLHYLGDVNPQMAAFNLSYPMGPHSKYENFVDTTIKNNYIVSDSNGYWN----- 203
D 124 AMHYFGDIDTYPHPANVT--AVDSAGHVKFTFAERKEQKINTAGCKTNEAFTDILK 181
QY 204 -----WKGANPEWIEGAAVAADQYDYPGVVNDTTKQMFVKAASQBYAD-KWRAEYTPV 256
D 182 NKDFNAMS-----KEYARGFATGKSIY-----YSHASHSHSDDWDYAAKVT-- 224
QY 257 TGKRLMEARVTAGYIH 273
D 225 ----LANSQKGTAGYIY 237

RESULT 2
US-08-725-518-2
; Sequence 2, Application US/08725518
; Patent No. 5851827
; GENERAL INFORMATION:
; APPLICANT: TITBALL DR, RICHARD W
; APPLICANT: WILLIAMSON DR, ETHEL D
; TITLE OF INVENTION: CLOSTRIDIUM PERFRINGENS VACCINES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE
; STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: UNITED STATES
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/725,518
; FILING DATE: 4-OCT-96
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/341538
; FILING DATE: 28-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9210717.6
; FILING DATE: 20-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9215655.3
; FILING DATE: 23-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO GB93/01039
; FILING DATE: 20-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: CRAWFORD, ARTHUR R
; REGISTRATION NUMBER: 25327
; REFERENCE/DOCKET NUMBER: 124-536
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 370 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-725-518-2

Query Match 14.5%; Score 220; DB 2; Length 370;
Best Local Similarity 26.5%; Pred. No. 6.8e-14;

```

```

Matches 68; Conservative 40; Mismatches 95; Indels 54; Gaps 10;
QY 46 EGINSHLWVNRADIMSRNTTIVNPNET-----ALLNWRADLENGIYSADYENPYDNS 101
D 6 DGTGTHAMIVTQGVSIENDLSKNEPESVRKNLEILKENMHELQLGSTPYDYNKAYD-- 63
QY 102 TYASHFYDDPTGTTTYP-----FAKHAKETG-----AKYFNLAGOAYQNQDMQOAFYYLGL 152
D 64 LYQDFHWDPTDNNFNSKDNSWYLAYSIDPTGESQIRKFSALARYEWQRGNYKQATFYLGE 123
QY 153 SLHYLGDVNPQMAAFNLSYPMGPHSKYENFVDTTIKNNYIVSDSNGYWN----- 203
D 124 AMHYFGDIDTYPHPANVT--AVDSAGHVKFTFAERKEQKINTAGCKTNEAFTDILK 181
QY 204 -----WKGANPEWIEGAAVAADQYDYPGVVNDTTKQMFVKAASQBYAD-KWRAEYTPV 256
D 182 NKDFNAMS-----KEYARGFATGKSIY-----YSHASHSHSDDWDYAAKVT-- 224
QY 257 TGKRLMEARVTAGYIH 273
D 225 ----LANSQKGTAGYIY 237

RESULT 3
US-09-510-322A-11
; Sequence 11, Application US/09510322A
; Patent No. 6391557
; GENERAL INFORMATION:
; APPLICANT: Fox Chase Cancer Center
; TITLE OF INVENTION: Nucleic Acid Encoding a Mismatch
; TITLE OF INVENTION: Endonuclease and Methods of Use Thereof
; FILE REFERENCE: 9503dna
; CURRENT APPLICATION NUMBER: US/09/510,322A
; CURRENT FILING DATE: 2000-02-22
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Plant
; US-09-510-322A-11

Query Match 6.3%; Score 96.5; DB 4; Length 276;
Best Local Similarity 21.4%; Pred. No. 0.14;
Matches 46; Conservative 31; Mismatches 67; Indels 71; Gaps 12;
QY 103 YASHFY----DPDTGTTTYPFAKHAKETGAKYFNLAGOAYQN-----QDM 143
D 54 WTSHLYYIDTDPQACSYEYSDCHDQGLKDMCVDG-AIQNFTSQLQHYGSGTSDRRYNN 112
QY 144 QOAFYYLGLSLHYLGDVNPQMAAFNFTN-----LSYPMGFH----- 179
D 113 TEALLFLS---HPMGDIIHQPH-VTFISDEGNTIDLRWYKXNLLHHVWDREIILTALK 168
QY 180 SKYENFVDTTIKNNYIVSDSNGYWN-----NWKGAN-----PEDWI-EGAAVAAKQDYPGV 228
D 169 ENYDKNLDLLQBDLEKNITNGLWHDDLSSTWTECDNLACPHKIASSESIKLACKWYKGVK 228
QY 229 NDTTKQMFVKAASQBYADKWRAEYTPVYTGRLME 263
D 229 SGET-----LSEEFYN----TRLPIVMKRIVQ 251

RESULT 4
US-08-491-954-4
; Sequence 4, Application US/08491954
; Patent No. 6096321
; GENERAL INFORMATION:
; APPLICANT: Girardeau, Jean-Pierre
; APPLICANT: Martin, Christine
; APPLICANT: Mechin, Marie-Claire
; APPLICANT: Der Vartanian, Maurice
; APPLICANT: Bousquet, Francois

```

TITLE OF INVENTION: SUB-UNIT OF CS31A PROTEIN CAPSULE
TITLE OF INVENTION: MODIFIED BY AT LEAST ONE HETEROLOGOUS PEPTIDE, CS31A
TITLE OF INVENTION: PROTEIN CAPSULE INCLUDING SUCH A SUB-UNIT, AND
TITLE OF INVENTION: MICROORGANISMS WHOSE OUTER MEMBRANE CARRIES SUCH
SUB-UNITS, AND PROCEDURE FOR OBTAINING AND UTILIZING SUCH
SUB-UNITS
NUMBER OF SEQUENCES: 115
CORRESPONDENCE ADDRESS:
ADDRESSEE: WEISER & ASSOCIATES
STREET: 230 South Fifteenth Street, Suite 500
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/491.954
FILING DATE: 16-FEB-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/01281
FILING DATE: 21-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 989.6264P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 274 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-491-954-4

Query Match 6.2%; Score 95; DB 3; Length 274;
Best Local Similarity 21.3%; Pred. No. 0.2;
Matches 64; Conservative 40; Mismatches 115; Indels 82; Gaps 13;
QY 1 MKKKVLAALAAVLAAPVQSVVPAQTNNSESP-----APILWSAEDKHNEGINS 50
DB 1 MKKTLIALA--VAVSAVSGSVMAADWTEGQPGDIIIGGEITSPSVKWLW--KTGEGUSS 56
QY 51 HLWIVNRAIDIMSRNTTIVNPNETALLNEWRADLENGIYSADYENPYDNDSTYASHFYDP 110
DB 57 ---FSNTTNEIVKRLNLSVPTDELFLAAKSDGKGVFVGNLTLPKTEMASDGSVITS 113
QY 111 DTGTTYIPFAXHAKETG-----AKYFN-----LAGQAYQ 139
DB 114 FTSNTAMDIADVKNKSGDNTLGTLSVPLSFGAAVTIFDGNITDSAVAHIITSGSAGTVFE 173
QY 140 N-----QDMQQAFFYGLS--LHYLGDVNPQPHANFTNLSYPMGFHSKYENFVDYTIKN 191
DB 174 GLVNPGRFTDQNIAYKNGLSKAEAMAGYVEKLMKPGKSSTSYS---GFH---NWDLSHP 226
QY 192 NYIVSDSNGYNWKGANPEDWTEGAAGAAQDYPGVNDTTKD-----WFKAAVSQBY 245
DB 227 NYTSADKASYLSY-----GSGVSAGST---LVWMLNKDVAGRLWEVAPVTITVIY 273
QY 246 A 246
DB 274 S 274

RESULT 5
US-08-282-197C-54

; Sequence 54, Application US/08282197C
; Patent No. 5871730
; GENERAL INFORMATION:
; APPLICANT: Brzezinski, Ryszard
; APPLICANT: Dery, Claude V
; APPLICANT: Beaulieu, Carole
; TITLE OF INVENTION: Thermostable Xylanase DNA, Protein and
; TITLE OF INVENTION: Methods of Use
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/282.197C
; FILING DATE: 29-JUL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Cimbala, Michele A
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 1050.0410000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; US-08-282-197C-54

Query Match 6.2%; Score 94; DB 2; Length 399;
Best Local Similarity 19.7%; Pred. No. 0.45;
Matches 51; Conservative 35; Mismatches 109; Indels 64; Gaps 10;
QY 19 QSVVPAQTNNSESPAPILWSAEDKHNEG---INSHLWIVNRAIDIMSRNTTIVNPNETA 75
DB 142 QLLAPAEKRGQMRGHTLIW-----HNQPSWLTNGNWRDSSLAVMKNHITVTMTHYKG 196
QY 76 LINWRA-----DLENGIYSADYEN---PYDNTSYASHFYDPDGTGTYTIPF----- 119
DB 197 KIVEMDVANECDSDSGNGLRSSIWRNVIGQDYLDFAFYAREADPDALLFYNDYNIEDLG 256
QY 120 -----AXHAKETGAKYFNLAGQAYQONQDMQAFVGLSLHYLGDVNPQMH----- 165
DB 257 PKNVAVFNMVISMKEKRGVP---IDGVGF-----QCHFGMSPEYLSAQDNKRYAEI 307
QY 166 ---AANFTNLSYPM-----GFHSKYENFVDYTIKNYIVSDSNGYNWKGANPEDWIEG 215
DB 308 GVIVSFTBEIDIRIPOSENPATAFQVQANNYKELMKI CLANPNCNTFVWVGFTDKYTWIPG 367
QY 216 AAVAAKQDYPGVNDTTKD 234
DB 368 T-----FPGYGNPLIYD 379

RESULT 6
US-09-252-991A-32033
; Sequence 32033, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136


```

RESULT 9
US-09-510-322A-10
; Sequence 10, Application US/09510322A
; Patent No. 6391557
; GENERAL INFORMATION:
; APPLICANT: Fox Chase Cancer Center
; TITLE OF INVENTION: Nucleic Acid Encoding a Mismatch
; FILE OF INVENTION: Endonuclease and Methods of Use Thereof
; FILE REFERENCE: 9503dna
; CURRENT APPLICATION NUMBER: US/09/510,322A
; CURRENT FILING DATE: 2000-02-22
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Plant
US-09-510-322A-10

Query Match          6.1%; Score 93; DB 4; Length 275;
Best Local Similarity 21.9%; Pred. No. 0.32;
Matches 46; Conservative 31; Mismatches 63; Indels 70; Gaps 14;

Qy 106 HFYD-PDTGTYTPAKHAK-EGAYFNLAGQYQNDMQAIFYLG----- 151
Db 60 HFIDTPDACSP-DYSRCDHPKGAEDMCVAG-AVHNVTTLQMHYRDGTSRRYNLTESL 117
Qy 152 -LSLHLYLGDVNOPMH-----AANFTNL---SYPMGFHSKYENFV----- 186
Db 118 LFSLHFMGDIHQPHVGTSGEGNTILNRFRHKSNIHVVWDREIILITAKAKYGGKOLD 177
Qy 187 -DTIKNNYIVSDNSGYW-----NWKGAN-----PEDWI-EGAAVAARQDYPGVVNDTTK 233
Db 178 AFQQLQNNF---TTGIWSDTSNGECDDLFCPCPKWASESISLACKWYGVTPGET- 232
Qy 234 DWFVKAUVSQYADKWAERAEVTPVTKRLME 263
Db 233 -----LSDEYFN-----SRMPIVMKRIAQ 251

RESULT 10
US-09-092-315-13
; Sequence 13, Application US/09092315
; Patent No. 6399337
; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; APPLICANT: Ge, Zhongming
; TITLE OF INVENTION: ALPHA-1, 3-FUCOSYLTRANSFERASE
; FILE REFERENCE: 07254/049001
; CURRENT APPLICATION NUMBER: US/09/092,315
; CURRENT FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: US 60/048,857
; EARLIER FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-092-315-13

Query Match          6.0%; Score 91.5; DB 4; Length 372;
Best Local Similarity 26.2%; Pred. No. 0.72;
Matches 44; Conservative 29; Mismatches 64; Indels 31; Gaps 10;

Qy 70 NPNETALNEWRADLENGIYSADYENPYDNYSTYASHFYD-----PDTG-----TTYIP 118
Db 164 HPNLCAVVNDESDPLKRGFASFVSNP---NAPIRNAFYDALNSTEPTGGGSRVNTLGY 220
Qy 119 FAKHAKETGAKY-FNLA-----GQAYQNDMQAIFYLGLSLHYLDVNOPMHAANFTNL 172
Db 221 NVKKNNEFLSQYKFNLCFPENTQGYGVTEKIIDAYFSHTIPIYW----GSPSVAKDFNPK 276

RESULT 11
US-09-092-315-1
; Sequence 1, Application US/09092315
; Patent No. 6399337
; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; APPLICANT: Ge, Zhongming
; TITLE OF INVENTION: ALPHA-1, 3-FUCOSYLTRANSFERASE
; FILE REFERENCE: 07254/049001
; CURRENT APPLICATION NUMBER: US/09/092,315
; CURRENT FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: US 60/048,857
; EARLIER FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-092-315-1

Query Match          6.0%; Score 91.5; DB 4; Length 464;
Best Local Similarity 26.2%; Pred. No. 1;
Matches 44; Conservative 29; Mismatches 64; Indels 31; Gaps 10;

Qy 70 NPNETALNEWRADLENGIYSADYENPYDNYSTYASHFYD-----PDTG-----TTYIP 118
Db 164 HPNLCAVVNDESDPLKRGFASFVSNP---NAPIRNAFYDALNSTEPTGGGSRVNTLGY 220
Qy 119 FAKHAKETGAKY-FNLA-----GQAYQNDMQAIFYLGLSLHYLDVNOPMHAANFTNL 172
Db 221 NVKKNNEFLSQYKFNLCFPENTQGYGVTEKIIDAYFSHTIPIYW----GSPSVAKDFNPK 276

RESULT 12
US-09-733-524A-1
; Sequence 1, Application US/09733524A
; Patent No. 6534298
; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; APPLICANT: Ge, Zhongming
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING ALPHA-1,3
; TITLE OF INVENTION: FUCOSYLTRANSFERASES AND EXPRESSION SYSTEMS FOR MAKING AND
; TITLE OF INVENTION: EXPRESSING THEM (amended)
; FILE REFERENCE: 07254-049002
; CURRENT APPLICATION NUMBER: US/09/733,524A
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 09/092,315
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: US 60/048,857
; PRIOR FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-733-524A-1

Query Match          6.0%; Score 91.5; DB 4; Length 464;
Best Local Similarity 26.2%; Pred. No. 1;
Matches 44; Conservative 29; Mismatches 64; Indels 31; Gaps 10;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2004, 14:46:13 ; Search time 47 Seconds
(without alignments)
1878.925 Million cell updates/sec

Title: PCT-US03-12556-2

Perfect score: 1522

Sequence: 1 MKKKVLAAMVLAAPVQS.....EAQRTAGYIHLWFDYVNR 282

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1288442 seqs, 313154207 residues

Total number of hits satisfying chosen parameters: 1288442

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1522	100.0	282	15	US-10-421-654-2
2	1257.5	82.6	397	15	Sequence 2, Appli
3	1232	80.9	283	15	Sequence 16747, A
4	1206.5	79.3	280	15	US-10-421-654-4
5	522	34.3	473	15	US-10-421-654-6
6	515	33.8	473	15	US-10-421-654-82
7	509	33.4	473	15	US-10-421-654-90
8	501	32.9	473	15	US-10-421-654-98
9	120	7.9	251	15	US-10-421-654-32
10	108.5	7.1	2140	12	US-10-282-122A-73670
11	102.5	6.7	872	12	US-10-424-599-239481
12	96.5	6.3	836	9	US-09-858-525A-10
13	96.5	6.3	871	9	US-09-858-525A-2
14	96	6.3	293	12	US-10-424-599-219446
15	93.5	6.1	479	9	US-09-733-524-17

Sequence 55192, A
Sequence 6, Appli
Sequence 28, Appli
Sequence 264682,
Sequence 13, Appli
Sequence 13, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 7, Appli
Sequence 7, Appli
Sequence 13784, A
Sequence 47689, A
Sequence 4171, Ap
Sequence 239478,
Sequence 120169,
Sequence 70739, A
Sequence 39, Appli
Sequence 2, Appli
Sequence 5, Appli
Sequence 269, App
Sequence 43305, A
Sequence 1, Appli
Sequence 4, Appli
Sequence 191038,
Sequence 77188, A
Sequence 20, Appli
Sequence 249, App
Sequence 211603,

16 93.5 6.1 787 12 US-10-282-122A-55192
17 93 6.1 664 13 US-10-067-385-6
18 2119 6.1 134 10 US-09-769-744A-28
19 92.5 6.1 372 12 US-10-424-599-264682
20 91.5 6.0 372 13 US-10-120-319-13
21 91.5 6.0 372 14 US-10-189-977-13
22 91.5 6.0 464 13 US-10-120-319-1
23 91.5 6.0 464 14 US-10-189-977-1
24 91.5 6.0 464 14 US-10-392-098-1
25 91.5 6.0 478 13 US-10-120-319-7
26 91.5 6.0 478 14 US-10-189-977-7
27 91.5 6.0 478 14 US-10-392-098-7
28 91.5 6.0 802 15 US-10-369-493-13784
29 90.5 5.9 584 12 US-10-282-122A-47689
30 90.5 5.9 934 15 US-10-369-493-4171
31 90 5.9 270 12 US-10-424-599-239478
32 90 5.9 328 16 US-10-437-963-120169
33 90 5.9 721 12 US-10-282-122A-70739
34 90 5.9 1365 16 US-10-383-930-39
35 89.5 5.9 636 9 US-09-843-007-2
36 89.5 5.9 636 16 US-10-705-195-5
37 89 5.8 218 14 US-10-174-693-269
38 89 5.8 866 12 US-10-282-122A-43305
39 88.5 5.8 501 9 US-09-733-524-1
40 88.5 5.8 517 15 US-10-381-596A-4
41 88.5 5.8 887 16 US-10-437-963-191038
42 88.5 5.8 894 12 US-10-282-122A-77188
43 88.5 5.8 1786 16 US-10-637-544-20
44 88 5.8 1659 16 US-10-474-776-249
45 87.5 5.7 195 12 US-10-424-599-211603

ALIGNMENTS

RESULT 1
US-10-421-654-2
; Sequence 2, Application US/10421654
; Publication No. US20040005604A1
; GENERAL INFORMATION:
; APPLICANT: Gramatikova, Svetlana
; APPLICANT: Hazlewood, Geoff
; APPLICANT: Lam, David E.
; APPLICANT: Barton, Nelson R.
; TITLE OF INVENTION: PHOSPHOLIPASES, NUCLEIC ACIDS ENCODING THEM AND
; TITLE OF INVENTION: METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-094001
; CURRENT APPLICATION NUMBER: US/10/421,654
; CURRENT FILING DATE: 2003-04-21
; PRIOR APPLICATION NUMBER: US 60/374,313
; PRIOR FILING DATE: 2002-04-19
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
US-10-421-654-2

Query Match 100.0%; Score 1522; DB 15; Length 282;
Best Local Similarity 100.0%; Pred. No. 1.9e-139;
Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKKKVLAAMVLAAPVQS...VFVFAQTNNSESPAPILRWSAEDKNEGINSHLWIVNRAID 60
DB 1 MKKKVLAAMVLAAPVQS...VFVFAQTNNSESPAPILRWSAEDKNEGINSHLWIVNRAID 60
QY 61 IMSRNTTIVNPNTALLNEWRADLENGIYSADYENPYNDSTYASHFYDPDTGTYIPFA 120

```
Db 1 IMSRNTTIVNPTALLNEWADLENGIYSADYENPYDNTSYASHFYDPDGTGYIPPA 120
Qy 121 KHAKETGAKYFNLAGQAYQNDQMOQAPFYLGSLHYLGDVNPQPMHAAFTNLSYPMGFHS 180
Db 121 KHAKETGAKYFNLAGQAYQNDQMOQAPFYLGSLHYLGDVNPQPMHAAFTNLSYPMGFHS 180
Qy 181 KYENFVDTIKNYIVSDSGYNNKNGANPEWIEGAAVAAKQDYPGVVNDTTKQWFKVAA 240
Db 181 KYENFVDTIKNYIVSDSGYNNKNGANPEWIEGAAVAAKQDYPGVVNDTTKQWFKVAA 240
Qy 241 VQSEYADKWRAEVTPVTGKRLMEARVVTAGYIHLWFDYTVNR 282
Db 241 VQSEYADKWRAEVTPVTGKRLMEARVVTAGYIHLWFDYTVNR 282

RESULT 2
US-10-369-493-16747
; Sequence 16747, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 16747
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Bacillus thuringiensis
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(397)
; OTHER INFORMATION: unsure at all xaa locations
US-10-369-493-16747

Query Match 82.6%; Score 1257.5; DB 15; Length 397;
Best Local Similarity 80.2%; Pred. No. 1.6e-113;
Matches 227; Conservative 23; Mismatches 32; Indels 1; Gaps 1;

Qy 1 MKKKVLALAAVMVALAAPVQSVVFAQTNNSESPAPIL-RWSAEDKHNEGINSHLWIVNRAI 59
Db 1 MKKKVLALAAITLVAPLQSVAFAHENDGQRFPGVIPRWSAEDKHKEGVNSHLWIVNRAI 60
Qy 60 DIMSRNTTIVNPTALLNEWADLENGIYSADYENPYDNTSYASHFYDPDGTGYIPF 119
Db 61 DIMSRNTTIVKQDRVALLNEWRTLENGIYAADYENPYDNTSYASHFYDPDNGKTYIPY 120
Qy 120 AKHAKETGAKYFNLAGQAYQNDQMOQAPFYLGSLHYLGDVNPQPMHAAFTNLSYPMGFH 179
Db 121 AKQAKETGAKYFKLAGEYQNDKNAFFYLGLSLHYLGDVNPQPMHAAFTNLSYPMQGFH 180
Qy 180 SKYENFVDTIKNYIVSDSGYNNKNGANPEWIEGAAVAAKQDYPGVVNDTTKQWFKVKA 239
Db 181 SKYENFVDTIKNYIVSDSGYNNKNGINPEWIEGAAVAAKQDYPGVVNDTTKQWFKVRA 240
Qy 240 AVSQEYADKWRAEVTPVTGKRLMEARVVTAGYIHLWFDYTVNR 282
Db 241 AVSQEYADKWRAEVTPVTGKRLMDAQRVTAGYIQLWFDYTVNR 283

RESULT 3
US-10-421-654-4
; Sequence 4, Application US/10421654
; Publication No. US20040005604A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Gramatikova, Svetlana
; APPLICANT: Hazlewood, Geoff
; APPLICANT: Lam, David E.
; APPLICANT: Barton, Nelson R.
; TITLE OF INVENTION: PHOSPHOLIPASES, NUCLEIC ACIDS ENCODING THEM AND
; TITLE OF INVENTION: METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-094001
; CURRENT APPLICATION NUMBER: US/10/421,654
; CURRENT FILING DATE: 2003-04-21
; PRIOR APPLICATION NUMBER: US 60/374,313
; PRIOR FILING DATE: 2002-04-19
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(24)
US-10-421-654-4

Query Match 80.9%; Score 1232; DB 15; Length 283;
Best Local Similarity 76.5%; Pred. No. 3e-111;
Matches 215; Conservative 35; Mismatches 31; Indels 0; Gaps 0;

Qy 1 MKKKVLALAAVMVALAAPVQSVVFAQTNNSESPAPILRWSAEDKHNEGINSHLWIVNRAID 60
Db 1 MKRKILATASVIALTAPIQSVAFAHENGHPPIALKWSAESIHNEGVSLSHLWIVNRAID 60
Qy 61 IMSRNTTIVNPTALLNEWADLENGIYSADYENPYDNTSYASHFYDPDGTGYIPPA 120
Db 61 IMSQNTTIVVQNETALLNEWRTLEKGIYSADYENPYDNTSYASHFYDPDGGKTYIPPA 120
Qy 121 KHAKETGAKYFNLAGQAYQNDQMOQAPFYLGSLHYLGDVNPQPMHAAFTNLSYPMGFHS 180
Db 121 KQAKQTGAKYFKLAGEYQNDKNAFFYLGLSLHYLGDVNPQPMHAAFTNLSHPFGFHS 180
Qy 181 KYENFVDTIKNYIVSDSGYNNKNGANPEWIEGAAVAAKQDYPGVVNDTTKQWFKVAAA 240
Db 181 KYENFVDTIKNYIVSDSGYNNQSNANPEWIEGAAVAAKADPPSI VNDKTKNWFVAAA 240
Qy 241 VQSEYADKWRAEVTPVTGKRLMEARVVTAGYIHLWFDYTVN 281
Db 241 VQSDSADKWRAEVTPITGKRLMEARVVTAGYIHLWFDYTVN 281

RESULT 4
US-10-421-654-6
; Sequence 6, Application US/10421654
; Publication No. US20040005604A1
; GENERAL INFORMATION:
; APPLICANT: Gramatikova, Svetlana
; APPLICANT: Hazlewood, Geoff
; APPLICANT: Lam, David E.
; APPLICANT: Barton, Nelson R.
; TITLE OF INVENTION: PHOSPHOLIPASES, NUCLEIC ACIDS ENCODING THEM AND
; TITLE OF INVENTION: METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-094001
; CURRENT APPLICATION NUMBER: US/10/421,654
; CURRENT FILING DATE: 2003-04-21
; PRIOR APPLICATION NUMBER: US 60/374,313
; PRIOR FILING DATE: 2002-04-19
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
```



```
RESULT 7
US-10-421-654-98
; Sequence 98, Application US/10421654
; Publication No. US20040005604A1
; GENERAL INFORMATION:
; APPLICANT: Gramatikova, Svetlana
; APPLICANT: Lam, David E.
; APPLICANT: Barton, Nelson R.
; TITLE OF INVENTION: PHOSPHOLIPASES, NUCLEIC ACIDS ENCODING THEM AND
; TITLE OF INVENTION: METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-094001
; CURRENT APPLICATION NUMBER: US/10/421,654
; PRIOR FILING DATE: 2003-04-21
; PRIOR FILING DATE: 2002-04-19
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 98
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Unknown
; OTHER INFORMATION: Obtained from an environmental sample
; NAME/KEY: SIGNAL
; LOCATION: (1)....(25)
US-10-421-654-98

Query Match 33.4%; Score 509; DB 15; Length 473;
Best Local Similarity 37.8%; Pred. No. 1.2e-40;
Matches 119; Conservative 43; Mismatches 109; Indels 44; Gaps 10;

Qy 1 MKKKVLALAAVVALAAPVQSV-----VFAQTNNSESPAPILWSAEDKHNEGIN 49
Db 1 MKKKLCTWALVTAISSGVVAIPTVASACGMEVMKQOEHK--RVKWSAEHPHSHES 58
Qy 50 SHLMVNRADIMSRN--TTVNPNETALLN--EWRADLENGIYSADYENPYDNST--- 102
Db 59 THLWIAQNAIQIMSRNQDKTQENELQFLNPPEYKELFERGLYDADYDFNDGGTGTIG 118
Qy 103 -----YASHFYDDPTGTYIIPAKHAK-----TGAKYFNLAGQAYQNDMQQAFY 149
Db 119 IDGLIRGCGKSHFYDPDKNY-----KGESEPTALSQGDYFKLAGEYFKKNDWKQAFY 174
Qy 150 LGLSLHYLDGVNQPMAAFTNLS--YPMGFHSKYENFVDTIKNNYIVSDSNGYNNWKGAN 208
Db 175 LGVATHYFTDATQPMHAANFTAVDRSAIKFHSAFEDYVTTIQEQPKVSDGEGKYNLVNSN 234
Qy 209 -PEDWIEGAAVAAKQDYPGVVNDTTKQWFKAAVQSEYADKWRABVTPTVGKRLMEAQRV 267
Db 235 DPQKWIHETARLAKVEIGNITNDVIKSHYNNKGNNA-----LWQEVMPAVQVSLEQAQRN 289
Qy 268 TAGYIHLWFDYVNR 282
Db 290 TAGFIHLWFKTYVGK 304

RESULT 8
US-10-421-654-32
; Sequence 32, Application US/10421654
; Publication No. US20040005604A1
; GENERAL INFORMATION:
; APPLICANT: Gramatikova, Svetlana
; APPLICANT: Lam, David E.
; APPLICANT: Barton, Nelson R.
; TITLE OF INVENTION: PHOSPHOLIPASES, NUCLEIC ACIDS ENCODING THEM AND
; TITLE OF INVENTION: METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-094001
; CURRENT APPLICATION NUMBER: US/10/421,654
; PRIOR FILING DATE: 2003-04-21
; PRIOR FILING DATE: 2002-04-19
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 106
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Unknown
; OTHER INFORMATION: Obtained from an environmental sample
; NAME/KEY: SIGNAL
; LOCATION: (1)....(30)
US-10-421-654-106

Query Match 32.9%; Score 501; DB 15; Length 473;
Best Local Similarity 36.5%; Pred. No. 7.4e-40;
Matches 115; Conservative 49; Mismatches 107; Indels 44; Gaps 10;

Qy 1 MKKKVLALAAVVALAAPVQSV-----VFAQTNNSESPAPILWSAEDKHNEGIN 49
Db 1 MKKKLCTWALVTAISSGVVAIPTVASACGMEVMKQOEHK--RVKWSAEHPHSHES 58
Qy 50 SHLMVNRADIMSRN--TTVNPNETALLN--EWRADLENGIYSADYENPYDNST--- 102
Db 59 THLWIAQNAIQIMSRNQDKTQENELQFLNPPEYKELFERGLYDADYDFNDGGTGTIG 118
Qy 103 -----YASHFYDDPTGTYIIPAKHAK-----TGAKYFNLAGQAYQNDMQQAFY 149
Db 119 IDGLIRGCGKSHFYDPDKNY-----KGESEPTALSQGDYFKLAGEYFKKNDWKQAFY 174
Qy 150 LGLSLHYLDGVNQPMAAFTNLS--YPMGFHSKYENFVDTIKNNYIVSDSNGYNNWKGAN 207
Db 175 LGVATHYFTDATQPMHAANFTAVDRSAIKFHSAFENYVTTVQTFEVKDDKGTYNLVNSD 234
Qy 208 NPEDWIEGAAVAAKQDYPGVVNDTTKQWFKAAVQSEYADKWRABVTPTVGKRLMEAQRV 267
Db 235 DPQKWIHETARLAKVEIGNITNDVIKSHYNNKGNNA-----LWQEVMPAVQVSLEQAQRN 289
Qy 268 TAGYIHLWFDYVNR 282
Db 290 TAGFIHLWFKTYVGK 304

RESULT 9
US-10-421-654-106
; Sequence 106, Application US/10421654
; Publication No. US20040005604A1
; GENERAL INFORMATION:
; APPLICANT: Gramatikova, Svetlana
; APPLICANT: Lam, David E.
; APPLICANT: Barton, Nelson R.
; TITLE OF INVENTION: PHOSPHOLIPASES, NUCLEIC ACIDS ENCODING THEM AND
; TITLE OF INVENTION: METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-094001
; CURRENT APPLICATION NUMBER: US/10/421,654
; CURRENT FILING DATE: 2003-04-21
; PRIOR FILING DATE: 2002-04-19
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 106
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Unknown
; OTHER INFORMATION: Obtained from an environmental sample
; NAME/KEY: SIGNAL
; LOCATION: (1)....(30)
US-10-421-654-106
```



```
RESULT 12
US-09-858-525A-10
; Sequence 10, Application US/09858525A
; Publication No. US20020199215A1
; GENERAL INFORMATION:
; APPLICANT: BOETS, Annemie
; APPLICANT: ARNAUT, Greta
; APPLICANT: VAN RIE, Jeroen
; APPLICANT: DAMME, Nicole
; TITLE OF INVENTION: No. US20020199215A1e1 Toxins
; FILE REFERENCE: 021565-077
; CURRENT APPLICATION NUMBER: US/09/858,525A
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: US 09/573,872
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 10
; TYPE: PRT
; LENGTH: 836
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ISPIA protein
; US-09-858-525A-10
```

```
Query Match
Best Local Similarity 6.3%; Score 96.5; DB 9; Length 836;
Matches 64; Conservative 48; Mismatches 140; Indels 53; Gaps 14;

QY 8 LAAMVALAAPVQSVVFAQT---NNSESPAPILRWSAEDKHNEGINSHLMIYNRAIDIM-- 62
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
267 VAAFPVNVVQSHSETVANEGSATNDGTHINGAESAYLANVRNNVGTGAIVETKPTTSFI 319
QY 63 SRNTIIVNPNETALLNEWRADLENGIYSADYENPY-----YDN-STYASHFYDPT--- 112
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
320 SFGVSVNYQHSETVANEGSATNDGTHINGAESAYLANVRNNVGTGAIVETKPTTSFI 379
QY 113 --GTTYIPPAKHAKETGAKYFNLAGQAYQNDMQQAFFYGLSLHYLGDVNPMPHAAFT 170
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
380 LDGTT-IGTIK-AKENTTALTILDPQSYPEKGN-----GIAINTMDDFNSRPIPLNKE 431
QY 171 NLSYPMGFHSHKYENFVDITIKNIYIVSDSNG-----YMNWKGANPEDWIEGAAV--- 218
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
432 QLNTYLSNKKPILLETQDVEGKYAIKDTNGNITTAGDMNGITDEISAKTASIIYDNGNQ 491
QY 219 ----AAQDYPGVVNDTTKDFWVKAASQEVADKWRAEVTPVTGKRLMEAQRTVAGYIHL 274
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
492 SEKRVAAKDYTN-PEDKTPNLSVKEALKLAYPD-----EIEBKDGLLFYNDQPIFEASVQS 546
QY 275 WFDTY 279
Db :|:
547 YVDEY 551
```

```
RESULT 13
US-09-858-525A-2
; Sequence 2, Application US/09858525A
; Publication No. US20020199215A1
; GENERAL INFORMATION:
; APPLICANT: BOETS, Annemie
; APPLICANT: ARNAUT, Greta
; APPLICANT: VAN RIE, Jeroen
; APPLICANT: DAMME, Nicole
; TITLE OF INVENTION: No. US20020199215A1e1 Toxins
; FILE REFERENCE: 021565-077
; CURRENT APPLICATION NUMBER: US/09/858,525A
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: US 09/573,872
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
```

```
; SEQ ID NO 2
; LENGTH: 871
; TYPE: PRT
; ORGANISM: Brevibacillus laterosporus
US-09-858-525A-2

Query Match
Best Local Similarity 6.3%; Score 96.5; DB 9; Length 871;
Matches 64; Conservative 48; Mismatches 140; Indels 53; Gaps 14;

QY 8 LAAMVALAAPVQSVVFAQT---NNSESPAPILRWSAEDKHNEGINSHLMIYNRAIDIM-- 62
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
302 VAAFPVNVVQSHSETVANEGSATNDGTHINGAESAYLANVRNNVGTGAIVETKPTTSFI 354
QY 63 SRNTIIVNPNETALLNEWRADLENGIYSADYENPY-----YDN-STYASHFYDPT--- 112
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
355 SFGVSVNYQHSETVANEGSATNDGTHINGAESAYLANVRNNVGTGAIVETKPTTSFI 414
QY 113 --GTTYIPPAKHAKETGAKYFNLAGQAYQNDMQQAFFYGLSLHYLGDVNPMPHAAFT 170
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
415 LDGTT-IGTIK-AKENTTALTILDPQSYPEKGN-----GIAINTMDDFNSRPIPLNKE 466
QY 171 NLSYPMGFHSHKYENFVDITIKNIYIVSDSNG-----YMNWKGANPEDWIEGAAV--- 218
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
467 QLNTYLSNKKPILLETQDVEGKYAIKDTNGNITTAGDMNGITDEISAKTASIIYDNGNQ 526
QY 219 ----AAQDYPGVVNDTTKDFWVKAASQEVADKWRAEVTPVTGKRLMEAQRTVAGYIHL 274
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
527 SEKRVAAKDYTN-PEDKTPNLSVKEALKLAYPD-----EIEBKDGLLFYNDQPIFEASVQS 581
QY 275 WFDTY 279
Db :|:
582 YVDEY 586

RESULT 14
US-10-424-599-219446
; Sequence 219446, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 219446
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(293)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_40186C.1.pep
US-10-424-599-219446

Query Match
Best Local Similarity 6.3%; Score 96; DB 12; Length 293;
Matches 47; Conservative 31; Mismatches 85; Indels 58; Gaps 9;

QY 104 ASHEYD-PDTGTYIIPFAKHAKETGAKYFNLAGQA-----YQNO-----DM 143
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
79 ALHFANTPDSVCNY-----KDTKDCVDKKTGKRCVAAITNYTNQLLEYGSDTSKYNL 134
QY 144 QQAFFYGLSLHYLGDVNPMPHAAFT-----NLSYPMG----- 177
Db :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
135 TQSLLEF---HFLGDVHQPLHCGFFSDKGGNEINVRWYRRKQNLHVVWDASIIETEVER 191
```


This page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2004, 14:38:52 ; Search time 17 Seconds
(without alignments)
1595.648 Million cell updates/sec

Title: PCT-US03-12556-2
Perfect score: 1522
Sequence: 1 MKKVLALAMVALAAPVQS.....EAQRTAGYIHLWFDYYNR 282

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1252.5	82.3	592	2 PS0197	phospholipase C (E
2	511.5	33.6	289	2 C43868	lecithinase - List
3	511	33.6	289	2 AF1100	phospholipase C (i
4	509	33.4	289	2 S20888	phospholipase C (E
5	314.5	20.7	178	2 T43165	probable phospholi
6	254	16.7	398	2 B30565	phospholipase C (E
7	231	15.2	398	2 JQ0366	phospholipase C (E
8	229	15.0	398	2 A30565	phospholipase C (E
9	227	14.9	398	2 B49231	phospholipase C, a
10	109.5	7.2	245	2 H97025	phospholipase C re
11	108.5	7.1	2140	2 F95074	serine proteinase,
12	107	7.0	2140	2 A97942	metalloproteinase,
13	103	6.8	625	2 G97076	alkaline phosphata
14	102.5	6.7	470	2 AF1972	hypothetical prote
15	102	6.7	562	2 S55516	glucan 1,3-beta-gl
16	102	6.7	2358	2 T39569	probable alpha-glu
17	99	6.5	273	2 D82862	conjugal transfer
18	98	6.4	472	2 AE0616	hypothetical prote
19	97.5	6.4	363	2 E90620	outer membrane pro
20	97	6.4	468	2 H64802	ybfM protein - Esc
21	97	6.4	653	2 H85620	partial fimbrial u
22	97	6.4	653	2 F90756	partial fimbrial u
23	96.5	6.3	532	2 D96966	beta-mannanase Man
24	95.5	6.3	418	1 A53888	thermolabile hemol
25	95.5	6.3	1283	2 S52500	oxyesterol-binding
26	95	6.2	441	2 EP0051	hypothetical prote
27	95	6.2	624	2 E64483	hypothetical prote
28	95	6.2	1599	2 S22737	glucosyltransferase
29	94.5	6.2	305	2 H86245	bifunctional nucle

RESULT 1

PS0197

phospholipase C (EC 3.1.4.3) precursor - Bacillus cereus

N:Alternate names: cereolysin AB; phosphatidylcholine-hydrolyzing phospholipase C

C:Species: Bacillus cereus

C>Date: 30-Sep-1991 #sequence revision 26-Apr-1996 #text change 15-Oct-1999

C:Accession: S18978; S0197; I40160; S01949; A32042; S03200; I40161

R:Kuzmin, N.P.; Gavrilenko, I.V.; Krukov, V.M.; Karpov, A.V.

submitted to the EMBL Data Library January 1992

A:Description: Fused genes phospholipase C and sphingomyelinase

A:Reference number: S18978

A:Accession: S18978

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-592 <K>

A:Cross-references: EMBL:X64140; NID:g39421; PIDN:CAA45501.1; PID:g39422

R:Ikeda, K.; Inoue, S.; Anasaki, C.; Teshima, K.; Ikezawa, H.

J. Biochem. 110, 88-95, 1991

A>Title: Kinetics of the hydrolysis of monodispersed and micellar phosphatidylcholines

A:Reference number: PS0197; MUID:92041771; PMID:1939031

A:Accession: PS0197

A:Molecule type: protein

A:Residues: 39-49 <IK>

R:Gavrilenko, I.V.; Baida, G.E.; Karpov, A.V.; Kuz'min, N.P.

Bioorg. Khim. 19, 133-138, 1993

A>Title: [Nucleotide sequence of phospholipase C and sphingomyelinase genes from Bacillus

A:Reference number: I40160; MUID:93249510; PMID:9387306

A:Accession: I40160

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-592 <RES>

A:Cross-references: EMBL:X64140; NID:g39421; PIDN:CAA45501.1; PID:g39422; EMBL:X64141; N

R:Johansen, T.; Haugli, F.B.; Ikezawa, H.; Little, C.

Nucleic Acids Res. 16, 10370, 1988

A>Title: Bacillus cereus strain SE-1: nucleotide sequence of the sphingomyelinase C gene

A:Reference number: S01949; MUID:89057484; PMID:2848222

A:Accession: S01949

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-283 <JOH>

A:Cross-references: EMBL:X12854; NID:g39444; PIDN:CAA31332.1; PID:g295765

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1988

R:Gilmore, M.S.; Cruz-Rodaz, A.L.; Leimeister-Waechter, M.; Kreft, J.; Goebel, W.

J. Bacteriol. 171, 744-753, 1989

A>Title: A Bacillus cereus cytolytic determinant, cereolysin AB, which comprises the pho

A:Reference number: A32042; MUID:89123149; PMID:2536680

A:Accession: A32042

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-8, 'G', 10-13, 'L', 15-30, 'ORFGVIP', 38-76, 'L', 78-119, 'Y', 121-211, 'D', 213-225, '

A:Cross-references: GB:M4149; NID:g142672; PIDN:AA91819.1; PID:g142673

R:Yamada, A.; Tsukagoshi, N.; Ueda, S.; Sasaki, T.; Makino, S.; Nakamura, S.; Little, C

Eur. J. Biochem. 175, 213-220, 1988
A>Title: Nucleotide sequence and expression in *Escherichia coli* of the gene coding for s
A;Reference number: S01130; MUID:88296483; PMID:2841128
A;Accession: S03200
A;Molecule type: DNA
A;Residues: 166-211, 'D', 213-225, 'A', 227-238, 'R', 240-281, 'N', 283 <YAM>
A;Cross-references: EMBL:X12711; NID:g39439; PIDN:CAA31213.1; PID:g575735
C;Genetics:
A;Gene: cerAB; plc
C;Superfamily: Bacillus cereus phospholipase C
C;Keywords: phosphoric diester hydrolase
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-38/Domain: propeptide #status predicted <PRO>
F;39-283/Product: phospholipase C #status predicted <MAT>

Query Match 82.3%; Score 1252.5; DB 2; Length 592;
Best Local Similarity 80.2%; Pred. No. 3.2e-91;
Matches 227; Conservative 23; Mismatches 32; Indels 1; Gaps 1;

Qy 1 MKKKVLAAAMVALAAPVQSVVFAQTNNSESPAPIL-RWSAEDKHNGINSHLMTVNRAI 59
Db 1 MKKKVLAAAITVVAPLQSVAFAPAHENGGSKIKIVHWSAEDKHKGVNSHLMTVNRAI 60

Qy 60 DIMSNTTIVNPNETALLNEWRADLENGIYSADYENPYDSTYASHFYDPTGTGYTYPF 119
Db 61 DIMSENTTLVKQDRVAQLNWRTELENGIYAADYENPYDSTPASHFYDPTDNGKTYTYPF 120

Qy 120 AKHAKETGAKYFNLAGQAYQNDMQQAFYGLSLHYHLDVNPQMAANFTNLSYPMGFH 179
Db 121 AKQAKETGAKYFKLAGESYKNDKMQAFYGLSLHYHLDVNPQMAANFTNLSYPMGFH 180

Qy 180 SKYENFVDTIKNNYIVSDSGYNNWKGANPEDWIEGAAVAAKQDYPGVNDDTKDMFVKA 239
Db 181 SKYENFVDTIKNNYIVSDSGYNNWKGANPEDWIEGAAVAAKQDYPGVNDDTKDMFVKA 240

Qy 240 AVSQYADKRAEAVTPVTKRLMEAQRTAGYIHLWFDTYNR 282
Db 241 AVSQYADKRAEAVTPVTKRLMEAQRTAGYIHLWFDTYGR 283

RESULT 2
C43868
lecithinase - *Listeria monocytogenes*
C;Species: *Listeria monocytogenes*
C;Date: 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 24-Sep-1998
R;Vazquez-Boland, J.A.; Kocks, C.; Dramsi, S.; Ohayon, H.; Geoffroy, C.; Mengaud, J.; Co
Infect. Immun. 60, 219-230, 1992
A>Title: Nucleotide sequence of the lecithinase operon of *Listeria monocytogenes* and pos
A;Reference number: A43868; MUID:92104678; PMID:1309513
A;Accession: C43868
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-289 <VAZ>
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:74437, NCBIIP:74462)
C;Superfamily: Bacillus cereus phospholipase C

Query Match 33.6%; Score 511.5; DB 2; Length 289;
Best Local Similarity 38.7%; Pred. No. 4.5e-33;
Matches 98; Conservative 52; Mismatches 96; Indels 7; Gaps 2;

Qy 24 AQTNNSESPAP-ILRWSAEDKHNGINSHLMTVNRADIMSRNTTIVNPNETALLNEWRA 82
Db 37 ARPHDIDSKPLKLSWSADNPTNTDVTNHYLWLFKQAEKILAKDVNHRANLNLKPKDK 96

Qy 83 DLENGIYSADYENPYDSTYASHFYDPTGTGYTYPFAKHAKETGAKYFNLAGQAYQND 142
Db 97 QTAQIYDADKHKNPYDYDTSLSHFNPDRDNTYLPGFANAKITGAKYFNQSVTDYREGK 156

Qy 143 MQQAFYGLSLHYHLDVNPQMAANFTNLSYPMGFHSHKYENFVDTIKNNYIVSDSGY 202
Db 157 FDTAFYKGLAIHYTDTISQPMHANNFTAISYPPGPHCAENYVDTIKHNYQATDWMVAK 216

Qy 203 NMKGANPEDWIEGAAVAAKQDYPGVNDDTKDMFVAAVSOEYADKRAEAVTPVTKRLM 262
Db 217 RFGSDDDKDWLYENAKRAKADYPIVNAKTKSVLGV-----NSEWKKDVTVEFTGARLR 270

Qy 263 EAQRTAGYIHLW 275
Db 271 DSQQTLAGFLFVW 283

RESULT 3
AF1100
phospholipase C [imported] - *Listeria monocytogenes* (strain EGD-e)
C;Species: *Listeria monocytogenes*
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AF1100
R;Glaser, P.; Prangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madieno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of *Listeria species*.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AF1100
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-289 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAD00732.1; PID:g16409570; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: plcB
C;Superfamily: Bacillus cereus phospholipase C

Query Match 33.6%; Score 511; DB 2; Length 289;
Best Local Similarity 37.1%; Pred. No. 5e-33;
Matches 106; Conservative 55; Mismatches 107; Indels 18; Gaps 5;

Qy 2 KKKVLAAAMVA-LAAPVQSVV-----FAQT----NNSESPAP-ILRWSAEDKHNGIN 49
Db 4 KKVVLGMLIASLVLPVTIKANACCDEYLOTPAAPHDIDSKLPHKLSWSADNPTNTDYN 63

Qy 50 SHLWIVNRADIMSRNTTIVNPNETALLNEWRADLENGIYSADYENPYDSTYASHFYD 109
Db 64 THYLFKQAEKILAKDVNHRANLNLKPKDKQIAQGIYDADHKNPYDSTSLSHFYN 123

Qy 110 PDTGTTTYPFAKHAKETGAKYFNLAGQAYQNDMQQAFYGLSLHYHLDVNPQMAANF 169
Db 124 PDRDNTYLPGFANAKITGAKYFNQSVTDYREGKFTAFYKGLAIHYTDTISQPMHANNF 183

Qy 170 TNLSPMGFHSKYENFVDTIKNNYIVSDSGYNNWKGANPEDWIEGAAVAAKQDYPGVN 229
Db 184 TAISYPPGPHCAENYVDTIKHNYQATDWMVAKRFGSDDDKDWLYENAKRAKADYFKLVN 243

Qy 230 DTTKDMFVAAVSOEYADKRAEAVTPVTKRLMEAQRTAGYIHLW 275
Db 244 AKTKSVLGV-----NSEWKKDVTVEFTGARLRDSQQTLAGFLFVW 283

RESULT 4
S20888
phospholipase C (EC 3.1.4.3) precursor - *Listeria monocytogenes* (fragment)
C;Species: *Listeria monocytogenes*
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 24-Sep-1999
C;Accession: S20888
R;Domann, E.; Wehland, J.; Rohde, M.; Pistor, S.; Hartl, M.; Goebel, W.; Leimeister-Wae
EMBO J. 11, 1981-1990, 1992
A;Title: A novel bacterial virulence gene in *Listeria monocytogenes* required for host ce
A;Reference number: S20887; MUID:92258410; PMID:1582425
A;Accession: S20888
A;Molecule type: DNA
A;Residues: 1-289 <DOM>
A;Cross-references: EMBL:X59723; NID:g44091; PIDN:CAA42408.1; PID:g444093

R:Tso, J.Y.; Siebel, C.
Infect. Immun. 57, 468-476, 1989
A>Title: Cloning and expression of the phospholipase C gene from Clostridium perfringens
A:Reference number: A30565; MUID:89108588; PMID:2536356
A:Accession: B30565
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-398 <Tso>
C:Keywords: phosphoric diester hydrolase

Query Match 16.7%; Score 254; DB 2; Length 398;
Best Local Similarity 26.9%; Pred. No. 1.5e-12;
Matches 73; Conservative 49; Mismatches 95; Indels 54; Gaps 10;

Qy 40 AEDKHNEGINSHLWVNRDAIDMSRNTTIIVNPNET-----ALLNEWRADLENGIYSAD 92
Db :
26 AWDGKKGDTGTSLIAEHGLSMLND--LSGNESQQVKDNIKILNEVLGDLKLSTYDP 82
Qy 93 YENPYDNSTYASHYPDPDTGTTY-----IPAKHAKETGAKYFNLGAOAYQNODM 143
Db :
83 YDPNAYD--LYQDHFPDPTGNNFTTDNSWIASYPIYDTSRSNVRKPFATLAKNEWEGNF 140
Qy 144 QOAPFYLGSLHYLDGVNOPMHAANFTNISYPMGFSKYENFVDITKNNIIVSDS---- 198
Db :
141 KEATFLQGQLHYLDGNTPYHASNTAVDSP--GHVKYTFVEERKDNTALNTSGNDTT 198
Qy 199 NGYNWKGANP-----EDWTIEGAVAARQDYPGVVNDTTKDFVFKAASQEYADKWRA 251
Db :
199 SGVYKEAMENPSFNKWMTONSIKYAKIA-----KDIYYSHSTMHSWDDW-- 243
Qy 252 EVTPVTGKRLLMEARQV-TAGYIHLWFDPTVVN 281
Db :
244 ---DYSGREAIKNSQVCTAGFLYRFNMVSN 271

RESULT 7
JQ0366
phospholipase C (EC 3.1.4.3) precursor - Clostridium perfringens
N:Alternate names: Clostridium oedematiens beta- and gamma-toxins; Clostridium welchii
C:Species: Clostridium perfringens
C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 15-Oct-1999
C:Accession: JQ0366
R:Saint-Joanis, B.; Garnier, T.; Cole, S.T.
Mol. Gen. Genet. 219, 453-460, 1989
A>Title: Gene cloning shows the alpha-toxin of Clostridium perfringens to contain both s
A:Reference number: JQ0366; MUID:90158507; PMID:2560137
A:Accession: JQ0366
A:Molecule type: DNA
A:Residues: 1-398 <SAI>
A:CROSS-references: GB:X17300; NID:g40619; PIDN:CAA35186.1; PID:g40620
A:Experimental source: strain 8-6
C:Comment: C. perfringens is a common pathogen of man and domestic animals and responsib
C:Comment: This enzyme has both sphingomyelinase and lecithinase activities.
C:Genetics:
A:Gene: plc
C:Keywords: phosphoric diester hydrolase
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-398/Product: phospholipase C #status predicted <MAT>

Query Match 15.2%; Score 231; DB 2; Length 398;
Best Local Similarity 28.4%; Pred. No. 1e-10;
Matches 86; Conservative 44; Mismatches 105; Indels 68; Gaps 16;

Qy 1 MKKKV---LALAAMV-ALAAPVQSUVFAQTNNSESAPILRWAEADKHNEGINSHLWTVN 56
Db :
1 MKRKICKALVCATLVTSLSWAGVSTKYVA-----W---DGKIDGTGFHAMIVT 44
Qy 57 RAIDI---MSRNTTIIVNPNETALLNEWRADLENGIYSADYENPYDNSTYASHFYDPDT 112
Db :
45 QGVSIENDMKNEPESVRKNLEILXDNMHLEQLGSTYPDYDKNAVY--LYQDHFWDPDT 102
Qy 113 GTTYIP-----PAKHAKETG-----AKYFNLGAOAYQNODMQQAFFYLGLSLHYLDGVNQP 163
Db :

Db 103 NNFPSKNSWYLAYSIPDTGSGQIRKFSALARYEWRQGNKYQATFYLGEMAHYFGDIDTP 162
Qy 164 MHAANFTNLSYPMGPHSKYENFVDTIKNNYIVSDSNGYNNKGANPEMDWIEGAQVAAKOD 223
Db 163 YHPANVT--AVDSAGHVKFETFAERKEQYKINTVGCCTN-----EDFY--ADILKNKD 212
Qy 224 YPGVNDTTKDW---FVKAASQBYA-----DKW--RAEVTPTVKRLMEARVTAQ 270
Db 213 F----NANSKEYARGFAKTGKSIYYSHASMSHSDWDYAAKVT-----LANSQKGTAG 262
Qy 271 YIH 273
Db 263 YIY 265

RESULT 8

A30565
phospholipase C (EC 3.1.4.3) precursor - Clostridium perfringens
N/Alternate names: alpha toxin
C/Species: Clostridium perfringens
C/Date: 23-Mar-1989 #sequence_revision 23-Mar-1989 #text_change 21-Jul-2000
C/Accession: A30565; A32418; A30564; S06996
R/TSO, J.Y.; Siebel, C.
Infect. Immun. 57, 468-476, 1989
A/Title: Cloning and expression of the phospholipase C gene from Clostridium perfringens
A/Reference number: A30565; MUID:89108588; PMID:2536356
A/Accession: A30565
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: DNA
A/Residues: 1-398 <TSO>
R/Okabe, A.; Shimizu, T.; Hayashi, H.
Biochem. Biophys. Res. Commun. 160, 33-39, 1989
A/Title: Cloning and sequencing of a phospholipase C gene of Clostridium perfringens.
A/Reference number: A32418; MUID:89228034; PMID:2540749
A/Accession: A32418
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-16,'S',18-398 <OKA>
A/Cross-references: GB:M24904; NID:g144889; PIDN:AAA23272.1; PID:g144889
R/Ticball, R.W.; Hunter, S.E.C.; Martin, K.L.; Morris, B.C.; Shuttleworth, A.D.; Rubidge
Mol. Microbiol. 3, 383-392, 1989
Infect. Immun. 57, 367-376, 1989
A/Title: Molecular cloning and nucleotide sequence of the alpha-toxin (phospholipase C)
A/Reference number: A30564; MUID:89108574; PMID:2536355
A/Accession: A30564
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: DNA
A/Residues: 1-16,'S',18-398 <TIT>
R/Leslie, D.; Fairweather, N.; Pickard, D.; Dougan, G.; Kehoe, M.
Mol. Microbiol. 3, 383-392, 1989
A/Title: Phospholipase C and haemolytic activities of Clostridium perfringens alpha-toxin
A/Reference number: S06996; MUID:8913290; PMID:2546005
A/Accession: S06996
A/Molecule type: DNA
A/Residues: 1-16,'S',18-398 <LES>
A/Cross-references: EMBL:X13608; NID:g40554; PIDN:CAA31943.1; PID:g40555
C/Genetics:
A/Genes: cpa
C/Keywords: phosphoric diester hydrolase
F/1-28/Domain: signal sequence #status predicted <SIG>
F/29-398/Product: phospholipase C #status predicted <MAT>

Query Match 15.0%; Score 229; DB 2; Length 398;
Best Local Similarity 25.8%; Pred. No. 1.5e-10;
Matches 79; Conservative 44; Mismatches 109; Indels 74; Gaps 13;

Qy 1 MKKKYLALAAWVALA----APQSVVFAQTNNSSPAPILRWSAEDKHNEGINSHLMTVN 56
Db 1 MKRKICKALICAAATLWAGASTKYA-----W---DGKIDGTGTHAMIVT 44
Qy 57 RAIDMSRNTTIVNPNET-----ALLNWRADLENGIYSADYENPYDNSTYASHFYDPT 112
Db 45 QGVSIENDLSKNEPESVRKNLEILKENMHQLQGSTPYDYNKAYD--LYODHFWDPDT 102

Qy 113 GTTYIP-----FAKHAKETG-----AKYFNLAGQAYQNDMQQAFYFGLSLHLGLGVDNQP 163
Db 103 DNNFSKNSWYLAYSIPDTGSGQIRKFSALARYEWRQGNKYQATFYLGEMAHYFGDIDTP 162
Qy 164 MHAANFTNLSYPMGPHSKYENFVDTIKNNYIVSDSNGYNNKGANPEMDWIEGAQVAAKOD 208
Db 163 YHPANVT--AVDSAGHVKFETFAERKEQYKINTAGCKTNEAFYTDILKNKDFNAWS--- 217
Qy 209 PEDWIEGAQVAAKODYPCGVNDTTKDMFVKAAVSOEYAD--KWRAEVTPVTVKRLMEARV 267
Db 218 -KEYARGFAKTGKSIY-----YSHASMSHSDWDYAAKVT-----LANSQKGTAG 259
Qy 268 TAGYIH 273
Db 260 TAGYIY 265

RESULT 9

B49231
phospholipase C, alpha-toxin - Clostridium perfringens
C/Species: Clostridium perfringens
C/Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Nov-1994
C/Accession: B49231
R/Katayama, S.; Matsushita, O.; Minami, J.; Mizobuchi, S.; Okabe, A.
Infect. Immun. 61, 457-463, 1993
A/Title: Comparison of the alpha-toxin genes of Clostridium perfringens type A and C strains
A/Reference number: B49231; MUID:93138764; PMID:8423073
A/Contents: type C, NCIB 10662
A/Accession: B49231
A/Status: preliminary
A/Molecule type: nucleic acid
A/Residues: 1-398 <KAT>
A/Note: sequence extracted from NCBI backbone (NCBIN:123169, NCBIPI:123171)

Query Match 14.9%; Score 227; DB 2; Length 398;
Best Local Similarity 26.7%; Pred. No. 2.1e-10;
Matches 81; Conservative 46; Mismatches 108; Indels 68; Gaps 15;

Qy 1 MKKKV----LALAAWVALAAPVQSVVFAQTNNSSPAPILRWSAEDKHNEGINSHLMTVN 56
Db 1 MKRKICKALICRTATSLWAGASTKYA-----W---DGKIDGTGTHAMIVT 44
Qy 57 RAIDMSRNTTIVNPNET-----ALLNWRADLENGIYSADYENPYDNSTYASHFYDPT 112
Db 45 QGVSIENDLSKNEPESVRKNLEILKENMHQLQGSTPYDYNKAYD--LYODHFWDPDT 102
Qy 113 GTTYIP-----FAKHAKETG-----AKYFNLAGQAYQNDMQQAFYFGLSLHLGLGVDNQP 163
Db 103 DNNFSKNSWYLAYSIPDTGSGQIRKFSALARYEWRQGNKYQATFYLGEMAHYFGDIDTP 162
Qy 164 MHAANFTNLSYPMGPHSKYENFVDTIKNNYIVSDSNGYNNKGANPEMDWIEGAQVAAKOD 223
Db 163 YHPANVT--AVDSAGHVKFETFAERKEQYKINTAGCKTN-----EDFY--ADILKNKD 212
Qy 224 YPGVNDTTKDW---FVKAASQBYA-----DKW--RAEVTPTVKRLMEARVTAQ 270
Db 213 F----NANSKEYARGFAKTGKSIYYSHASMSHSDWDYAAKVT-----LANSQKGTAG 262
Qy 271 YIH 273
Db 263 YIY 265

RESULT 10

H97025
phospholipase C related protein [imported] - Clostridium acetobutylicum
C/Species: Clostridium acetobutylicum
C/Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C/Accession: H97025
R/Nolling, J.; Brston, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum

Qy	143	-----MQAPFFYGL--SLHYLGDVNQPMHAANF-----TNLSYPMGF	176
Dd	179	EGISSVERAQVQPMNMNHARKEIGVEAIDYLSINAPP-GKPFDRGMWISNIDTGDY	237
Qy	179	HSKYENFVDTRIK-----NNYIVSD-----SNGYWNWKGANPDWIEGAATAAA	220
Dd	238	RHKAMRIDDDAKASMRFKEDLKGTDQKWYLWSDKIPIHAFNYINGGKITVEKYDDG----	292
Qy	221	KQDY-----PGVV--NDTTKD	234
Dd	293	-RDYFDPHGMHIAGILAGNDTEQD	315
 RESULT 12 A97942			
C:Species: Streptococcus pneumoniae			
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001			
C:Accession: A97942			
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff			
e, R.; LeBlanc, D.J.; Lee, N.; Lefkowitz, E.D.; Lu, J.; Matsushima, P.; McArthur,			
y, P.; Sun, P.M.C.; Winkler, M.E.			
J. Bacteriol. 183, 5709-5717, 2001			
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskun			
A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.			
A:Reference number: A97872; MUID:21429245; PMID:11544234			
A:Accession: A97942			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-2144 <KUR>			
A:Cross-references: GB:AE007317; PIDN:AAK99365.1; PID:g15458138; GSPDB:GN00174			
C:Genetics:			
A:Gene: prtA			
C:Keywords: hydrolase; serine proteinase			
Query Match 7.0%; Score 107; DB 2; Length 2144;			
Best Local Similarity 23.1%; Pred. No. 5.7;			
Matches 76; Conservative 36; Mismatches 113; Indels 104; Gaps 19;			
Qy	1	MKKKV---LALAMVALAAPQSVFPQTNNSSPAPI--LRWSAEDKHREGINSHLWI-	54
Dd	1	MKKSITVLSTTAAILAAVAPNEVLADTSSSEDALSISDKVKVVDKETENKEKHDIH	60
Qy	55	--VNRAIDMSRNTTV-----NP-----NETALNEWRADLENGIYSADY----	93
Dd	61	NALETSQDTPEEKTTTIEEEKVSKPKVIDTKTSNEEAKTBENSQSOGDHDTDSFVNKN	120
Qy	94	-ENP-YDYNSTYASHFVPDPTGTITYIPFAKHAKETGAKY-----FNLAGAYQ----NQD	142
Dd	121	TENPKKEDKVVIYAEPFKESGEKAIGLSNLKNTKLVLYTYDIRFN--GSAIETTDPNLD	178
Qy	143	-----MQOAFYLGH--SLHYLDVQNPMHAANF-----TNLS	173
Dd	179	KIQIEGISIERAQVQPMNMNHARKEIGVEAIDYLSINAPP-GKPFDRGMWISNID	237
Qy	174	YPMGFSKYENFVDTIK-----NNYIVSD-----SNGYWNWKGANPDWIEG	215
Dd	238	TGTDYRHKAMRIODDDAKASMRFKEDLKGTDQKWYLWSDKIPIHAFNYINGGKITVEKYDDG	297
Qy	216	AAVAARKQDY-----PGVV--NDTTKD	234
Dd	298	-RDYFDPHGMHIAGILAGNDTEQD	320
 RESULT 13 G97076			
alkaline phosphatase superfamily protein [imported] - Clostridium acetobutylicum			
C:Species: Clostridium acetobutylicum			
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001			
C:Accession: G97076			
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson,			
R.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.			

Search completed: July 27, 2004, 14:47:48
Job time : 18 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2004, 23:41:59 ; Search time 3697 Seconds
(without alignments)
9953.540 Million cell updates/sec

Title: PCT-US03-12556-1
Perfect score: 849
Sequence: 1 atgaaaagaagattagc.....atagctatgtaaatcgctaa 849

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rtd.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	559.2	65.9	1675	1	BTY16268	Y16268 Bacillus th
2	556	65.5	2245	1	BACCRAB	M24149 Bacillus ce
3	556	65.5	2349	1	BQGENE	X61451 B. cereus ge
4	556	65.5	2700	1	BCPLCSMC	X12854 B. cereus g
5	554	65.3	2201	1	BCFUSE	X61440 B. cereus fu
6	552.8	65.1	290214	1	AE017026	AE017026 Bacillus
7	552.8	65.1	301289	1	AE017000	AE017000 Bacillus
8	546.4	64.4	2130	1	AY195600	AY195600 Bacillus
9	255.6	30.1	1622	1	BCPLC	X12711 B. cereus g
10	189.4	22.3	280	1	BACHEMOL	M35411 B. cereus he
11	176.8	20.8	825	1	AY150839	AY150839 Listeria
12	154.4	18.2	870	6	AX413749	AX413749 Sequence
13	154.4	18.2	870	6	AX415655	AX415655 Sequence
14	154.4	18.2	5648	1	LISACTLDH	M82881 Listeria mo
15	154.4	18.2	160050	1	AL591974	AL591974 Listeria
16	154.4	18.2	349880	6	AX638811	AX638811 Sequence
17	152.8	18.0	3031	1	LMACTPLC	X59723 L.monocytog
18	148.8	17.5	758	1	AY150836	AY150836 Listeria
19	148.2	17.5	758	1	AY150830	AY150830 Listeria
20	148.2	17.5	758	1	AY150840	AY150840 Listeria
21	148.2	17.5	758	1	AY150841	AY150841 Listeria
22	145.6	17.1	758	1	AY150837	AY150837 Listeria
23	144.6	17.0	758	1	AY150835	AY150835 Listeria
24	144	17.0	758	1	AY150833	AY150833 Listeria
25	144	17.0	758	1	AY150834	AY150834 Listeria
26	137.4	16.2	758	1	AY150838	AY150838 Listeria
27	118.2	13.9	3417	1	LIV249805	AJ249805 Listeria
28	117.4	13.8	4097	1	LIU19035	U19035 Listeria iv
29	117.2	13.8	740	1	AY150832	AY150832 Listeria
30	64.2	7.6	701	1	AY150831	AY150831 Listeria
31	57.4	6.8	4562	6	AX416781	AX416781 Sequence
32	57.2	6.7	1237	1	AF072123	AF072123 Clostridi
33	56.8	6.7	250029	3	AE014820	AE014820 Plasmodi
34	55	6.5	125623	3	AC115599	AC115599 Dictyoste
35	55	6.5	183357	2	EX569779	EX569779 Danio rer
36	54.2	6.4	340552	3	PFA929354	AL929354 Plasmodi
37	54	6.4	1234	1	AB061869	AB061869 Clostridi
38	53.4	6.3	156323	2	BX296532	BX296532 Danio rer
39	52.6	6.2	33321	3	AC116986	AC116986 Dictyoste
40	51	6.0	110000	3	AC116984_3	Continuation (4 of
41	50.8	6.0	84618	5	AL591365	AL591365 Zebrafish
42	50.6	6.0	110000	2	PFMAL13_24	Continuation (25 o
43	50.2	5.9	192578	5	AL732455	AL732455 Zebrafish
44	49.8	5.9	256879	3	AC116982	AC116982 Dictyoste
45	49.4	5.8	49144	2	AC131380	AC131380 Strongylo

ALIGNMENTS

RESULT 1
BTY16268
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

BTY16268 1675 bp DNA linear BCT 02-OCT-1998
Bacillus thuringiensis plca gene and partial plcB gene
Y16268
Y16268.1 GI:2815226
phospholipase C; plcA gene; plcB gene; sphingomyelinase.
Bacillus thuringiensis
Bacillus thuringiensis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus
cereus group.

1
Lougren,A., Carlson,C.R., Eskils,K. and Kolsto,A.B.
Localization of putative virulence genes on a physical map of the
bacillus thuringiensis subsp. gelechiae chromosome

JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
Source

Curr. Microbiol. 37 (4), 245-250 (1998)
9732531
2 (bases 1 to 1675)
Lovgren,A.
Direct Submission
Submitted (21-JAN-1998) A. Lovgren, Dept of Microbiology, Stockholm
University, S-106 91 Stockholm 106 91, SWEDEN
Location/Qualifiers
1. 1675
/organism="Bacillus thuringiensis"
/mol_type="genomic DNA"
/strain="Bt13"
/sub_species="gelechiaie"
/db_xref="taxon:1428"
/clone="PLC9:51"
484..1335
/gene="plcA"
484..1335
/gene="plcA"
/codon_start=1
/transl_table=11
/product="phosphatidyl-degrading phospholipase C"
/protein_id="CAA76148.1"
/db_xref="GI:2815227"
/db_xref="GOA:O52864"
/db_xref="SPTREMBL:O52864"
/translation="MKKKVLAALATLVAFLQSVAFHENDGGORFGVIRWASBDK
HKEGVNHLWLVNRAIDMSRNTLVKQDRVALLNWRTELENIYAADYENPYDNS
TFASHFYDPNGKTYIPAKOKSETGAKYFKLAGESYKMKQKQAFYFLGSLHYLGD
VNPQMAANFNLSYPOGFHSKYENFVDITKNDYKVTGNGYKMKWNTNPDWHLGAA
VVAQDVAGIVNDNTQFWRAVSVQEVADKWRAEVTPMTGKRLMDAQRTVAGYIQLW
FDYTGNR"
1529..1675
/gene="plcB"
1529..1675
/gene="plcB"
/codon_start=1
/transl_table=11
/product="sphingomyelinase"
/protein_id="CAA76149.1"
/db_xref="GI:2815228"
/db_xref="SPTREMBL:O52865"
/translation="MTHNVYMLSTNLYPNWQTERADLIGAADYIKNDVVDVILNEVPD
NSASD"

ORIGIN

Query Match 65.9%; Score 559.2; DB 1; Length 1675;
Best Local Similarity 79.3%; Pred. No. 2.5e-105;
Matches 676; Conservative 0; Mismatches 173; Indels 3; Gaps 1;

QY 1 ATGAAAAGAAAGTATTAGCACTAGCAGCTATGGTGTCTTAGCTGCGCCAGTTCARAGT 60
DB 484 ATGAAAAGAAAGTACTTGTCTTAGCGGAGCTATTACATTAGTTGCTCCATTACAAAGT 543
QY 61 GTAGTATTGTCACAAACAAATAATAGTG---AAAGTCTCGCACCCGATTTTAAGATGGTCA 117
DB 544 GTTGCTATTGCTCATGAATATGATGGGGAGAGAGATTGGAGTTATTCGGCGCTGGTCT 603
QY 118 GCTGAGGATAGCATATAATAGAGGGGATTAATCTCTCATTTGTGGATTGTAATCGTGCATAT 177
DB 604 GCTGAAGATAAACAATAAAGAGCGCGTGAATCTCATTTATGATGTTGTAACCGTGCATAT 663
QY 178 GACATCATGCTCGTAATACACAGATTGCTGATCCGATCGAATGAACCTGATTTAATATGAG 237
DB 664 GATATTATGCTCGTAATACACACTTGTAAACAAAGATCGAGTTGCACTATTAAATGAA 723
QY 238 TGGCGTGTGATTTAGAAAATGGTATTATCTGCTGATTACGAGAACTCTTATTATGAT 297
DB 724 TGGCGTACTGAGTAGAGAACGGTATTATGCTGCTGACTATGAATACCTTATTATGAT 783
QY 298 AATAGTACATATGCTTCTCATCTTTATGATCCGGATACCTGGAACAACATATATTCCTTTT 357
DB 784 AATAGCACATTTGCTTCACTTTCTATGACCTGCAATGGGAAAACCTTATATTCCTGAT 843

QY 358 GCGAAACATGCAAAAGAAAACAGCGCAAAATATTTTAACTTCTGCTGCTCAAGCATACCAA 417
DB 844 GCAAGCAGGCAAGGAACTGGAGCTAAATATTTTAAATAGCTGGTGAAGCTTACAAA 903
QY 418 AATCAAGATATGCAAGCAATCTTCTTACTTAGATATTCGCTTCAATTTATTTAGGAGAT 477
DB 904 AATCAAGATATGCAAGCAATCTTCTTACTTAGATATTCGCTTCAATTTATTTAGGAGAT 963
QY 478 GTGAATCAGCAATGCAATGCAAGCAATCTTCTTACTTAGATATTCGCTTCAATTTATTTAGGAGAT 537
DB 964 GTAAACCAACCGATGCAATGCAAGCAATCTTCTTACTTAGATATTCGCTTCAATTTATTTAGGAGAT 1023
QY 538 TCTAAATACGAAAATTTTGTGATACCAATAAAAATAAATATATTTTCTAGATAGCAAT 597
DB 1024 TCTAAATATGAAAATTTTGTGATACCAATAAAAATAAATATATTTTCTAGATAGCAAT 1083
QY 598 GGATATTGGAATTTGGAAGGAGCAACCCAGAGATTTGGATTTGAAGGAGCAGCGTAGCA 657
DB 1084 GGATATTGGAATTTGGAAGGAGCAACCCAGAGATTTGGATTTGAAGGAGCAGCGTAGCA 1143
QY 658 GCTAAACAGATTTCTGCGTGTGACGATACGACAAAGATTTGGTGTGTAAGCA 717
DB 1144 GCGAAACAGATTTCTGCGTGTGACGATACGACAAAGATTTGGTGTGTAAGCA 1203
QY 718 GCCGTATCTCAAGAAATATGCAAGATTAATGCGTGTGCGGAAGTAAACACCGTGTGACAGGAAAG 777
DB 1204 GCTGTATCACAAGAAATATGCAAGATTAATGCGTGTGCGGAAGTAAACACCGTGTGACAGGAAAG 1263
QY 778 CGTTAATGGAAGCGCGGTTACAGCTGGTGTATATTTCAATTTGTTGTTGATACGAT 837
DB 1264 CGTTAATGGAAGCGCGGTTACAGCTGGTGTATATTTCAATTTGTTGTTGATACGAT 1323
QY 838 GTAAATCGCTAA 849
DB 1324 GGAATCGTTAA 1335

RESULT 2
BACERAB
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
FEATURES
source

BACERAB
Bacillus cereus
M24149
24149.1
GI:142672
cerA gene; cerB gene.
Bacillus cereus
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.
1 (bases 1 to 2245)
Gilmore, M.S., Cruz-Rodriguez, A.L., Leimeister-Wachter, M., Kreft, J. and Goebel, W.
A Bacillus cereus cytolytic determinant, cereolysin AB, which comprises the phospholipase C and sphingomyelinase genes: nucleotide sequence and genetic linkage
J. Bacteriol. 171 (2), 744-753 (1989)
89123149
2536680
Original source text: Bacillus cereus DNA.
Location/Qualifiers
1. 2245
/organism="Bacillus cereus"
/mol_type="genomic DNA"
/db_xref="taxon:1396"
201..1052
/gene="cerA"
201..1052
/gene="cerA"
/codon_start=1
/transl_table=11
/protein_id="AAA91819.1"
/db_xref="GI:142673"
/translation="MKKKVLAALATLVAFLQSVAFHENDGGORFGVIRWASBDK"

HKEGVNSHLWVNRDAIDMSRNTTLVKQDRVALLNWEWTELENGIYAADYENPPYDNS
TFQPHFDPDNGKTYIPYAKQAKETGAKYFKLAGESYKNDKMQKQAFYGLSLHYLGD
VNQPMHAANFTNLSYPOGFHSHKYENFVDTIKDNYKVTGNGYNNWKGTPNPDWTHGAA
VVAQDYAGIVNDNTKDFVRAAVSQEVADKRAEVPMTGKRLMDAQRVTAGYIQLW
FDTYGNR"

gene

1129..2130

CDS

/gene="cerB"

1129..2130

/gene="cerB"

/codon_start=1

/transl_table=11

/protein_id="AAA91820.1"

/db_xref="GI:142674"

/translation="MKGKLLKGVLSLGVGLGALYSGTSAQAEASTNQNDTLKVMTHNV
YMLSTNLPVNGQTERLDFGAADYIKNDVILNEVFNDSASRLNKLKKEYPNQT
AVLGRSSSEVDKLTGNTSSSTPEDGVAIISKWPIAEKIQYVFAKGGCPGNLKNKGF
VYTKIKNDRFVHVTGLQAEDESMCGKTSASVTRNOLKEIQDFIKNKTNNNEYVL
IGDMVNVKINAEENNDSYASMEKTLNASVPSYTGHTATWDATNLSAKYNPEDSLA
EYLDYIIASKDHANPSYIENKVLQPKSPQNTWTFWPKNIRIMWILIIIIQVEATISMK"

ORIGIN

Query Match 65.5%; Score 556; DB 1; Length 2245;

Best Local Similarity 79.1%; Pred. No. 1.1e-104;

Matches 674; Conservative 0; Mismatches 175; Indels 3; Gaps 1;

Qy	1	ATGAAAAAGAAAGTATTAGCACTAGCAGCTATGGTGTCTTTAGCTGCGCCAGTTCAAAGT	60
Db	201	ATGAAAAAGAAAGTACTTGTCTTAGCGCAGCTATTACATTAGTTGCTCCATTACAAAGT	260
Qy	61	GTAGTATTGTCACAAACAATAATAGTG---AAAGTCCTGCACCGATTTTAAGATGGTCA	117
Db	261	GTTCGATTGTCTCATGAAATGATGGGGACAGAGATTGGAGTTATTTCGCGCTGCT	320
Qy	118	GCTGAGGATAGCATAATGAGGGGATTAACTCTCATTTGTGGATTGTAATCGTCAATT	177
Db	321	GCTGAGATTAACATTAAGAGCGCTGAATCTCTCATTTATGATTTGTAACCGTCAATT	380
Qy	178	GACATCATGTCTCGTAATACACAGTATGTGAATCCGAATGAACATGCATTATTAATGAG	237
Db	381	GATATTATGTCTCGTAATACACACTTGTAAACAAAGATCGAGTTGCATATTAAATGAA	440
Qy	238	TGGCGTGTGATTAGAAAATGGTATTATTCTCTGCTGATTACGAGAACTCTTATTATGAT	297
Db	441	TGGCGTGTGATTAGAAAATGGTATTATTCTCTGCTGCTGATGATGAATACTTATTATGAT	500
Qy	298	AATAGTACATGTCTCTCACATTTATGATCCGATACGGAACAAATATATATCTCTTTT	357
Db	501	AATAGCATTGCTTTCACATTTCTATGACCTGACATGGGAAACTTATTATCCGTAT	560
Qy	358	GCGAAACATGCAAAAAGAACAGCGCAAAATATTTTAACTTGTGTTGTTCAAGCATACCAA	417
Db	561	GCAAGCAGGCAAGGAAACTGGAGCTAAATATTTTAAATAGCTGTTGAGTCTTTACAAA	620
Qy	418	AATCAAGATATGCAGCAAGCATCTCTTACTTAGGATTTATGCTTCAATTTTAGAGAT	477
Db	621	AATAAGATATGAACAAGCATCTCTTATTTAGGATTTATCTCTTCAATTTTAGGGAT	680
Qy	478	GTGATACGCCAATCATGCGCAACATTTAGCAATCTTCTTATCCCAATGGGTTTCCAT	537
Db	681	GTAACCAACCAATGCATGCGGCAACCTTTAGCAACCTTTCGTATCCCAAGGATTTCCAT	740
Qy	538	TCTAAATACGAAAATTTTGTGATACAAATAAAAAATAACTATATTATTTTTCAGATGCAAT	597
Db	741	TCTAAATATGAAAATCTTTGTAGATACGATTAAGATTAATTAAGTAACCGATCGAAT	800
Qy	598	GGATATTGGAATTTGAAAGGAGCAAAACCCAGAGATTGGATTGAAGAGCAGCGGTAGCA	657
Db	801	GGGATTGGAATCGAAAGGTACGAATCCAGAAGATTGATTCATGAGCGGCGCAGTAGTT	860
Qy	658	GCTAAACAAGATTATCTTGGCGCTGTGACGATACGACAAAGATTGCTTTGTAAGAACA	717
Db	861	GCGAAAACAAGATTACGCTGGCATTGTAAATGATAATACGAAAGATTGGTTGCTAAGACA	920

Qy	718	CGCGTATCTCAAGAAATATGCAAGATAAATGGCGTGGGAAGTAAACACCGGTGACAGGAAG	777
Db	921	GCTGTATCACAGAATATGCAAGATAAATGGCGTGGGAAGTAAACACCGGTGACAGGAAG	980
Qy	778	CGTTTAAATGGAAGCGGCGGTTTACAGCTGTTTATATTTTGTGTTTGTATACGTTAT	837
Db	981	CGTTTAAATGGAAGCGGCGGTTTACAGCTGTTTATATTTTGTGTTTGTATACGTTAT	1040
Qy	838	GTAATATCGCTAA 849	
Db	1041	GGAAATCGTTAA 1052	
RESULT 3			
BCGENE			
LOCUS	B.cereus genes plc and sph.	2349 bp	DNA linear BCT 26-JUN-1997
DEFINITION	B.cereus genes plc and sph.		
ACCESSION	X64141		
VERSION	X64141.1 GI:312905		
KEYWORDS	phospholipase C; PLC gene; sph gene; sphingomyelinase.		
SOURCE	Bacillus cereus		
ORGANISM	Bacillus cereus		
REFERENCE	1 (bases 1 to 2349)		
AUTHORS	Gavrilenko, I.V., Baida, G.E., Karpov, A.V. and Kuz'min, N.P.		
TITLE	Nucleotide sequence of phospholipase C and sphingomyelinase genes from Bacillus cereus BKM-B164		
JOURNAL	Bioorg. Khim. 19 (1), 133-138 (1993)		
MEDLINE	93249510		
PUBMED	8387306		
REFERENCE	2 (bases 1 to 2349)		
AUTHORS	Kuzmin, N.P.		
TITLE	Direct Submission		
JOURNAL	Submitted (14-JAN-1992) N.P. Kuzmin, Inet. of Biochemistry & Physiology of Microorganisms AS USSR, Pustchino, MOSCOW Region, 142292, USSR		
COMMENT	See also x64140.		
FEATURES	Location/Qualifiers		
source	1..2349		
	/organism="Bacillus cereus"		
	/mol_type="genomic DNA"		
	/strain="VKM 164"		
	/db_xref="taxon:1396"		
	/clone="pHR3"		
RBS	41..45		
gene	53..904		
CDS	53..904		
	/gene="plc"		
	/gene="plc"		
	/EC_number="3.1.4.3"		
	/codon_start=1		
	/evidence=experimental		
	/transl_table=11		
	/product="phospholipase c"		
	/protein_id="CAA45502.1"		
	/db_xref="GI:312906"		
	/db_xref="GOA:P09598"		
	/db_xref="SWISS-PROT:P09598"		
	/translation="MKKVLALAAITVVAPLQVAFPAHENDGSGSKIIVHRWSAEDK HKEGVNSHLWVNRDAIDMSRNTTLVKQDRVALLNWEWTELENGIYAADYENPPYDNS TFQPHFDPDNGKTYIPYAKQAKETGAKYFKLAGESYKNDKMQKQAFYGLSLHYLGD VNQPMHAANFTNLSYPOGFHSHKYENFVDTIKDNYKVTGNGYNNWKGTPNPDWTHGAA VVAQDYAGIVNDNTKDFVRAAVSQEVADKRAEVPMTGKRLMDAQRVTAGYIQLW FDTYGNR"		
RBS	965..968		
gene	978..1979		
CDS	978..1979		
	/gene="sph"		
	/gene="sph"		
	/EC_number="3.1.4.12"		
	/codon_start=1		

Db 833 CGATTAATGATGCACAAAGCTGTTACTGCTGGATACATTCAGCTTTGGTTGTGATACGTAC 892

Qy 838 GTAAATCGCTAA 849

Db 893 GGAGATCGTTAA 904

RESULT 4

BCPLCSMC 2700 bp DNA linear BCT 26-JUN-1997

LOCUS B. cereus genes for phospholipase C and sphingomyelinase C.

DEFINITION X12854 M20194

ACCESSION X12854.1 GI:39444

VERSION phospholipase C; PLC gene; smc gene; sphingomyelinase.

KEYWORDS Bacillus cereus

SOURCE Bacillus cereus

ORGANISM Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.

REFERENCE 1 (bases 1 to 2050)

AUTHORS Johansen, T., Holm, T., Guddal, P.H., Sletten, K., Haugli, P.B. and Little, C.

TITLE Cloning and sequencing of the gene encoding the phosphatidylcholine-preferring phospholipase C of Bacillus cereus

JOURNAL Gene 65 (2), 293-304 (1988)

MEDLINE 88313678

PUBMED 3137122

REFERENCE 2 (bases 1441 to 2700)

AUTHORS Johansen, T., Haugli, P.B., Ikezawa, H. and Little, C.

TITLE Bacillus cereus strain SE-1: nucleotide sequence of the sphingomyelinase C gene

JOURNAL Nucleic Acids Res. 16 (21), 10370 (1988)

MEDLINE 89057484

PUBMED 2848222

REFERENCE 3 (bases 1 to 2700)

AUTHORS Johansen, T.

TITLE Direct Submission

JOURNAL Submitted (12-SEP-1988) Johansen T., Institute of Medical Biology, Cell Biology, University of Tromsø, 9001 Tromsø, Norway

COMMENT On Jul 24, 2002 this sequence version replaced gi:143336. Data kindly reviewed (11-JAN-1989) by Johansen T.

FEATURES

source Location/Qualifiers

1..2700 /organism="Bacillus cereus"

/mol_type="genomic DNA"

/strain="SE-1"

/db_xref="taxon:1396"

1..305 /note="ORF1"

291..340 /note="put. rho-independent terminator"

368..445 /note="intergenic palindromic sequence"

453..458 /note="pot. rRNA binding site"

476..481 /codon_start=1

545..554 /transl_table=11

/product="PLC"

/protein_id="CAA31332.1"

/db_xref="GI:295765"

/db_xref="GOA:P09598"

/db_xref="SWISS-PROT:P09598"

/translation="MKKVLALAAITVVAPLQSVAFHENDGGSKI KIVHRWSAEDK HKEGVNSHLWIVNRAIDIMSRNTLVKQDRVAQLNWEKTELENGIYAADYENPYDNS TFASHFYDNGKTYIPAKOKAKETGKAFYKLAGEYKWKDMKQAFYVLGSLHYLGD VNOPHAAFNFLSYPOGFHSKYENFDTIKDNYKVTGNGYWKWKGTPREWHGAA VVAKQDYSIGVINDNTKQWFWKAAVSQETADKWRAEVTPMTGKRLMDAQRVTAGYIQLW FDTYDGR"

562..633 /note="put. signal peptide (PLC)"

676..1410

misc_feature

terminator

misc_feature

-35_signal

-10_signal

RBS

CDS

sig_peptide

mat_peptide

/evidence=experimental

/transl_table=11

/product="sphingomyelin phosphodiesterase"

/protein_id="CAA45503.1"

/db_xref="GI:580755"

/db_xref="GOA:P09598"

/db_xref="SWISS-PROT:P09598"

/translation="MKKLLGVLSFGIGLVLYGGSSVQADTSTDONTLKUMTHNV YMLSTNLYPNNGQSRADLIGIADYIKQDVVILNEVFDNSADRLGNLKETPNQT AVLGRSNGNDKVTIGYSSTPREDGVAVIVSKWPIVEIKQYVFAKCGPDNLSNKG FVTIKKNDREPVHVIQTHLQAEQSMCGTSPASVTRNQLKEIQDFIKNQIPIINDEYVL PGGDMYVKNINAEKNSDEYASMEKTLHASIPSVTGHATWDAATNSIAKYNFDDSPA EYLDYIASKDHANPSPIENKVLQPKSQWTVTSWLKXYTYDDYSDHVPVAATISMK"

ORIGIN

Query Match 65.5%; Score 556; DB 1; Length 2349;

Best Local Similarity 79.1%; Pred. No. 1.1e-104;

Matches 674; Conservative 0; Mismatches 175; Indels 3; Gaps 1;

Qy 1 ATGAAAAAGAAAGTATTAGCACTAGCAGCTATGCTTGGTTAGCTGGCCAGTTCAGAGT 60

Db 53 ATGAAAAAGAAAGTATTAGCACTAGCAGCTATGCTTGGTTAGCTGGCCAGTTCAGAGT 112

Qy 61 GTAGTATTGGCAACAAATAATAGTGAAGTCTCTGACCGGATTTTAA---GATGGTCA 117

Db 113 GTTGCAATTGCTCATGAATAATGATGGGAGTAAATAAATAATAGTTCACCGCTGGTCT 172

Qy 118 GCTGAGGATAGCATATGAGGGGATTAATCTCTCATTTGTGATGTTAAATCGTCAATT 177

Db 173 GCTGAAGATAAACAATAAGAGGTTGTAATCTCATTTATGATGTTAAACCGTCCGATT 232

Qy 178 GACATCATGCTCGTAATACACAGTATGTAATCGAATGAACTGCATTATTAATGAG 237

Db 233 GATATATGCTCGCAATACACACTTGTGTAACAAAGATCGAGTTGCACAAATTAATGAA 292

Qy 238 TGGCGTGTGATTAGTAAATGATTTATTTCTGCTGATTAGAGATCTCTTATTATGAT 297

Db 293 TGGCGTACGAGTTAGAACCGTATTATGCTGCTGCTACTGAAATCTTTATTATGAT 352

Qy 298 AATAGTACATATGCTTCTCATTTTATGATCCGGATACCTGGAACAACATATATCTTTT 357

Db 353 AATAGCACATTTGCTTCAATTTCTATGATCCAGCAATGGAACATATATATCTTTT 412

Qy 358 GCGAATATGCAAAAGAACGCGCAAAATATTTTAACTTGTGCTGATCAAGCATACCAA 417

Db 413 GCAAGACGCAAAAGAACGCGCTAAATATTTTAAATTTAGTCTGGTGAATCATACAA 472

Qy 418 AATCAAGATATGCAAGCAATCTTCTACTTAGGATTAATCGCTTCAATTTATGAGAT 477

Db 473 AATAAGATATGAACAAGCAATCTTCTATTAGGATTAATCTTCAATTTATGAGAT 532

Qy 478 GTGAATCAGCAATGATGAGCAAACTTTACGAATCTTCTTATCCATGCTTCCAT 537

Db 533 GTAAACCAACCGATGATGCGGCAAACTTTTACAAATCTTTTATATCCACAGGATTTCCAT 592

Qy 538 TCTAATATGCAAAATTTTGTGTATACATAAATAAATAAATACTATTTTTCAGATGCAAT 597

Db 593 TCTAATATGCAAACTTTTGTGTATGATGATTAAGATTAATTAAGTACCGATGCAAT 652

Qy 598 GGATATTGGAATTTGAAAGAGCAAAACCCAGAGATTTGATTTGAAGAGCGGTTAGCA 657

Db 653 GGATATTGGAATTTGAAAGAGTACAAATCCAGAGAGTGGATTCATGAGCGCGTAGTA 712

Qy 658 GCTAACAAGATTTATCTGCGCTTTGTAACGATACGCAAAAGATTTGTTTGTAAAGCA 717

Db 713 GCGAAAACAAGTTATCTTGGAAATTTGTAATGATTAATACGAAGATTTGTTTGTAAAGCA 772

Qy 718 GCGGTATCTCAAGAATATGAGATAAATGCGGTGCGGAAGTAAACACCGGTGACAGAAAG 777

Db 773 GCTGTGTCACAAGAAATATGAGATAAATGCGCGCTGNAAGTTACCAACATGACAGGTAAG 832

Qy 778 CGTTTAAATGAAAGCGCGGTTACAGCTGGTTATTTCAATTTGTTGGTTGATAGTAT 837

```
/product="PLC"
/note="mature PLC exoenzyme"
1413..1465
/note="pot. rho-independent terminator"
1418..1423
1441..1446
1471..1479
/note="pot. rRNA binding site"
1487..2488
/codon_start=1
/transl_table=11
/product="SMase"
/protein_id="CAA31333.1"
/db_xref="GI:580761"
/db_xref="GOA:P09599"
/translation="MKGKLLKGLVLSFGIGLVYGGSSVQADSTSDQNTLTKVMTNHY
YMLSTNLYPNQGSORADLIGAADYIKNQDVILNEVFNDNSADRLGNLKEYPNOT
AVLGRSNGEWDKTLGYSVSSSTPEDGGVAIVSKPIVEKIQYVFAKGGCPDNLNKGK
VYTKIKGDRFVHVTGHLQADSCMGKTSKSPVRLNOLKEIQQDFKGNKIPNDRIYVL
FGDMNVNKNIAENNSDEYASMFKTLHSPSYTGHTATWDATNTAKNYFPDSPA
EYLDYITASKDHANPSPFIENKLPQSPQWTVTSWLKYYTDDYSDHYPAATISMK"
1487..1567
/note="put. signal peptide (SMase)"
1568..2485
/product="SMase"
/note="mature SMase exoenzyme"
2493..2528
/note="pot. rho-independent terminator"

ORIGIN

Query Match      65.5%; Score 556; DB 1; Length 2700;
Best Local Similarity 79.1%; Pred. No. 1.1e-104;
Matches 674; Conservative 0; Mismatches 175; Indels 3; Gaps 1;

QY  1 ATGAAAAGAAGTAGTACACTAGCAGCATGTTGCTTTAGCTGCGCCAGTTCAAGT 60
    |||||
DB  562 ATGAAAAGAAGTAGTACACTAGCAGCATGTTGCTTTAGCTGCGCCAGTTCAAGT 621
    |||||

QY  61 GTAGTATTGTCACAAACAAATAAGTAGTGAAGTCTCTGCACCGCATTTTAA---GATGGTCA 117
    |||||
DB  622 GTTGCATTGCTCATGAANAATGATGGGGAGTAAATAAATAAGTAGTCCACCGCTGGTCT 681
    |||||

QY  118 GCTGAGGATAAGCATAATGAGGGGATTAATCTCTCAATTTGTGATTTGTAATGTCGCAAT 177
    |||||
DB  682 GCTGAAGATAAACATAAAGAAGTGTAAATCTCATTTATGATTTGTAATGTAACCGTCCGATT 741
    |||||

QY  178 GACATCATGCTCGTAATACACGAGTGTGAATCCGAATGAACTGCAATTAATGAG 237
    |||||
DB  742 GATATTATGCTCGCAATACACACTTGTAAACAAAGATCGAGTTGCACAATTAATGAA 801
    |||||

QY  238 TGGCGTGTGATTAGAAAATGGTATTATTCTGCTGATTAGAGAACTCTTATTATGAT 297
    |||||
DB  802 TGGCGTACGGAGTAGAGAACGGTATTATTGCTGCTGACTATGAANAATCTTATTATGAT 861
    |||||

QY  298 AATAGTACATATGCTTCTCATCTTTATGATCGGGATCTGGAACCAACATATATTCCTTT 357
    |||||
DB  862 AATAGCACAATTTGCTTTCATCTTATGATCCAGCAATGGAACCAACATATATTCATTT 921
    |||||

QY  358 GCGAACATGCAAGAAACAGGCGCAAAATATTTTAACCTTGTGCTGCAAGCATACCAA 417
    |||||
DB  922 GCAAGCAGGCAAGAAAGAACTGGCGCTAAATATTTTAAATAGCTGGTGAATCATACAAA 981
    |||||

QY  418 AATCAAGATATGACAGACAGCTTCTTCTACTTAGGATTATGCTTCAATTTAGGAGAT 477
    |||||
DB  982 AATAAAGATATGAACAAGACAGCTTCTTCTATTAGGATTATCTCTCATTTATTAGGAT 1041
    |||||

QY  478 GTGAATCAGCCCAATGCATCGACAACTTTAGCAATCTTCTTATCCAAATGGGTTTCCAT 537
    |||||
DB  1042 GTAAACCAACCGATGCATCGGCAAACTTTACAAATCTTTCATATCCCAAGGATTCCAT 1101
    |||||

QY  538 TCTAAATACGAAAATTTTGTGATACATAAATAAATAAATCTATATTGTTTCAGATAGCAAT 597
    |||||
```

```
Db 1102 TCTAAATATGAAACTTTTGTAGATACGATAAAGATAATTTAAAGTACCGATGGAAT 1161
QY 598 GGATATTGGAATTTGGAAGAGGACAAACCCAGAGAATTGGATTGGAAGGACGCGTAGCA 657
    |||||
Db 1162 GGATATTGGAATTTGGAAGAGGATCAAAATCCAGAAGAGTGGATTTCATGGAGCGCAGTAGTA 1221
    |||||
QY 658 GCTAAACCAAGATTATCTCGCGTGTGTAACGATACGACAAAAGATTGGTTGTTAAAGCA 717
    |||||
Db 1222 GCGAAACCAAGATTACTCTGGAATTTGTAAATGATTAATACGAAGATTGGTTGTTAAAGCA 1281
    |||||
QY 718 GCCGTATCTCAAGAAATATGCAGATAAATGGCGTGGGAAGTAAACACCGGTGACAGGAAAG 777
    |||||
Db 1282 GCTGTGTCACAGAATATGCAGATAAATGGCGCGTGAAGTTACACCAATGACAGGTAG 1341
    |||||
QY 778 CGTTTAATGGAAGCGCAGCGCTTACAGCTGGTTATATTTCATTTGTTGTTGATACGTAT 837
    |||||
Db 1342 CGATTAAATGGATGCACACGTTTACTCTGATACATTCAGCTTGGTTGATACGTATAC 1401
    |||||
QY 838 GTAATATCGCTAA 849
    |||||
Db 1402 GGAGATCGTTAA 1413
    |||||

RESULT 5
BCFUSE BCFUSE B.cereus fuse genes plc and sph (=h-cerAB) linear BCT 17-JUN-1993
LOCUS X64140
DEFINITION X64140.1 GI:39421
ACCESSION X64140.1
VERSION X64140.1
KEYWORDS cereolysin AB, fuse genes.
SOURCE Bacillus cereus
ORGANISM Bacillus cereus
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.

REFERENCE
AUTHORS Gavrilenko, I.V., Baida, G.E., Karpov, A.V. and Kuz'min, N.P.
TITLE Nucleotide sequence of phospholipase C and sphingomyelinase genes from Bacillus cereus BKM-B164
JOURNAL Bioorg. Khim. 19 (1), 133-138 (1993)
MEDLINE 93249510
PUBMED 8387306
REFERENCE 2 (bases 1 to 2201)
AUTHORS Kuzmin, N.P.
TITLE Direct Submission
SUBMITTED (14-JAN-1992) N.P. Kuzmin, Inst. of Biochemistry & Physiology of Microorganisms AS USSR, Pustchino, MOSCOW Region, 142292, USSR
JOURNAL
COMMENT See also X64141.
FEATURES
location/Qualifiers
source 1..2201
/organism="Bacillus cereus"
/mol_type="genomic DNA"
/strain="VKM 164"
/db_xref="taxon:1396"
/clones="pf27f"
41..45
/note="SD h-cerAB"
53..1831
/gene="h-cerAB"
53..1831
/gene="h-cerAB"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="hybrid cereolysin AB"
/protein_id="CAA45501.1"
/db_xref="GI:39422"
/db_xref="GOA:P09598"
/db_xref="GOA:P09599"
/db_xref="SWISS-PROT:P09598"
/db_xref="SWISS-PROT:P09599"
/translation="MKKVLALAAATITVAPLQSVAFAHENDGSGSKIIVIRWSAEDK
HKEGVNSHLWITVNRADIMSRNTTLVKQDRVAQLNWRTELENGIYAADYENPYDYS
```



```
Qy 418 AATCAAGATATGACGCAAGCAATCTTCTTCTACTTAGGATTATCGCTTCATTATTTAGGAGAT 477
Db 601 AATAAGATATGAAACAAGCATTCTTCTATTATTAGGATTATCTTCTTCTATTATCTAGGGAT 660
Qy 478 GTGATCAGCAATGATGACGCAAACTTTTACGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 537
Db 661 GTAAACCAACCGATGATGACGCAAACTTTTACGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 720
Qy 538 TCTAAATACGAAATTTTGTGATACAATAAAAAATACTATATTTTGTGATACGATCAAT 597
Db 721 TCTAAATACGAAATTTTGTGATACAATAAAAAATACTATATTTTGTGATACGATCAAT 780
Qy 598 GGATATTCGAATTTGAAAGGACAAACCCAGAAAGATTGATTTGAAGGAGCAGCGGTAGCA 657
Db 781 GGATATTCGAATTTGAAAGGAGTACGAATCCAGAAAGATTGATTTCAATGAGCGGCGAGTAGTT 840
Qy 658 GCTAAACAAAGATTATCTCTGGCTTGTGACGATACGACAAAGATTGTTTGTGTAAGCA 717
Db 841 GCGAAACAAAGATTATCTCTGGCTTGTGACGATACGACAAAGATTGTTTGTGTAAGCA 900
Qy 718 GCCGTATCTCAAGATATGCAATAAATGGCGTGGGAAAGTAACACCGGTGACAGGAAAG 777
Db 901 GCTGTATCACAAGATATGCAATAAATGGCGTGGGAAAGTAACACCGGTGACAGGTAAG 960
Qy 778 CGTTTAAATGGAAGCCGACGCGTTTACAGCTGGTTATATTCATTTTGTGTTGATACGTAT 837
Db 961 CGTTTAAATGGAAGCCGACGCGTTTACAGCTGGTTATATTCATTTTGTGTTGATACGTAT 1020
Qy 838 GTAAATCGCTAA 849
Db 1021 GGAGATCGTTAA 1032
```

```
RESULT 9
BCPLC BCPLC 1622 bp DNA linear BCT 26-JUN-1997
LOCUS B. cereus gene for sphingomyelinase (EC 3.1.4.3).
```

```
DEFINITION X12711
ACCESSION X12711.1 GI:39439
VERSION 1
KEYWORDS phospholipase; phospholipase C; serine carboxypeptidase; sphingomyelinase.
```

```
SOURCE Bacillus cereus
ORGANISM Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.
```

```
REFERENCE 1 (bases 1 to 1622)
AUTHORS Yamada,A., Teukagoshi,N., Uchida,S., Sasaki,T., Makino,S., Nakamura,S., Little,C., Tomita,M. and Ikezawa,H.
TITLE Nucleotide sequence and expression in Escherichia coli of the gene coding for sphingomyelinase of Bacillus cereus
JOURNAL Eur. J. Biochem. 175 (2), 213-220 (1988)
MEDLINE 88296483
PUBMED 2841128
```

```
COMMENT Data kindly reviewed (25/4/89) by Teukagoshi N.
FEATURES
source
1..1622
Location/Qualifiers
/organism="Bacillus cereus"
/mol_type="genomic DNA"
/strain="IAM1208"
/db_xref="taxon:1396"
/clone_lib="pUC18."
<1..358
/codon_start=2
/transl_table=11
/product="phosphatidylcholine-hydrolyzing phospholipase C"
/protein_id="CAA31213.1"
/db_xref="GI:757735"
/db_xref="GOA:P09598"
/db_xref="SWISS-PROT:P09598"
/translation="HAANFTNLSYPOGFHFKYENFVDIKNYKVTDGNGYWNWKGITN
PEDWTHGAAVAKQDYAGIVNDNTKDFVRAAVSQEYADKWRAEVTPTMTGKRLMDAQR
VTAGYIQLWFPTYGNR"
```

CDS

```
promoter 363..368
/note="-.35 region"
repeat_unit 366..381
/note="inverted repeat A"
promoter 386..392
/note="-.10 region"
repeat_unit 388..404
/note="inverted repeat A'"
RBS 422..426
/note="pot. ribosome binding site"
CDS 435..1436
/codon_start=1
/transl_table=11
/product="presphingomyelinase (AA -27 to 306)"
/protein_id="CAA31213.1"
/db_xref="GI:580758"
/db_xref="GOA:P11889"
/db_xref="SWISS-PROT:P11889"
/translation="WKGLKGLVSLGVLGALYSGTSAQAEASTNQNDTLKVMTHNV
YMLSTNLYPNWGQTERADLIGAADYIKNDVVILNEVPDNASDRLLGNLKKYENQIT
AVLGRSGSEWDKRTIGNYSSSTPEDGGVAIVSRWPIAEIKIYVFAKGGCPDNLNKG
VYTKIKNDREHVFVTHLQAEQSMCGKTSFASVTRNQLKEIQDFIKNNIPNNEYVL
IGDMNVNKKINAEKNDSEYASMEFKTLNASVPSYTGHTATWDATNSIAKYNFPDSPA
EYLDYIIASKDHANPSYIENKVLQPKSPQWTVTSFQKYTYNDISDDYIPEVATISMK"
435..515
/note="signal peptide (AA -27 to -1)"
516..1433
/product="sphingomyelinase (AA 1 - 306)"
1441..1453
/note="inverted repeat B"
repeat_unit 1459..1471
/note="inverted repeat B'"
ORIGIN
```

```
Query Match 30.1%; Score 255.6; DB 1; Length 1622;
Best Local Similarity 82.1%; Pred. No. 1e-42; Mismatches 0; Gaps 0;
Matches 294; Conservative 0; Indels 0; Gaps 0;
Qy 492 GCATGCGAGCAAACTTTACGAATCTTCTTCTTCAATGGGTTTCCATTTCTAAATAGCAAAA 551
Db 1 GCATGCGGCAAACTTTACGAACCTTTCGTATCCACAAGGATTCCTATTCTAAATATGAAA 60
Qy 552 TTTTGTGATACATAAAAAATACTATATTTGTTTCAGATAGCAATGATATTTGGAATG 611
Db 61 CTTTGTAGATACGATAAAAGATAATATAAAGTAACCGATGGAATGGTATTTGGAACGTG 120
Qy 612 GAAAGGACCAACCAAGAGATTGGATTGAAGGAGCAGCGGTAGCAGTAAACAAGATTGA 671
Db 121 GAAAGGTCAGATCCAGAAAGATTGGATTTCATGGAGCGGCGAGTAGTTGCGAAACAAGATT 180
Qy 672 TCCTGGCGTTTGTGAACGATACGACAAAAGATTGGTTTGTAAAGCAGCGCGTATCTCAAGA 731
Db 181 CGCTGGCATTGTAATGATATACGAAGATTGGTTTCGTAAGAGCAGCTGATACAGA 240
Qy 732 ATATGCGAGATAAATGGCGTGGGAGTAACACCGGTGACAGAAAGCGTTTAAATGGAAGC 791
Db 241 ATATGCGAGATAAATGGCGCGCTGGAAGTTACACCAATGACAGTAAGCGTTTAAATGGAAGC 300
Qy 792 GCAGCGGTTACAGCTGGTTATATTTCTTTGTTTGTATCGTATGATGTAATCGCTAA 849
Db 301 ACAACGTTTACTGCTGGATACATTTCAGCTTTGGTTTGTATACGGAATTCGTTAA 358
```

RESULT 10

```
BACHEMOL BACHEMOL 280 bp DNA linear BCT 26-APR-1993
LOCUS B. cereus hemolysin gene, partial cds.
DEFINITION B. cereus hemolysin gene, partial cds.
ACCESSION M35411
VERSION M35411.1 GI:143048
KEYWORDS hemolysin.
SOURCE Bacillus cereus
ORGANISM Bacillus cereus
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus
```



```
Db 798 AGTAAGCGCTGCTACTCGAAGAGACTA 825
|||||
RESULT 12
AX413749
LOCUS AX413749 870 bp DNA linear PAT 02-SEP-2002
DEFINITION Sequence 740 from Patent WO0228891.
ACCESSION AX413749
VERSION AX413749.1 GI:21446206
KEYWORDS
SOURCE Listeria monocytogenes EGD-e
ORGANISM Listeria monocytogenes EGD-e
REFERENCE 1
AUTHORS Kunst, F. and Glaser, P.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 740 11-APR-2002;
INSTITUT PASTEUR (FR); CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES
source
Location/Qualifiers
1..870
/organism="Listeria monocytogenes EGD-e"
/mol_type="unassigned DNA"
/db_xref="taxon:169963"
ORIGIN
Query Match 18.2%; Score 154.4; DB 6; Length 870;
Best Local Similarity 53.1%; Pred. No. 8.2e-22;
Matches 329; Conservative 0; Mismatches 291; Indels 0; Gaps 0;
Qy 98 CACCGATTTTAAAGATGGTCAGCTGAGGATAAGCAATAATGAGGGGATTAACCTCTCATTTGT 157
|||
Db 140 CACATAAACTTAGTTGGTCCGGGATAACCGACAAATACTGACGTAAATAGCGACTATT 199
|||
Qy 158 GGATTGTAATCGTGCAATTGACATCATGCTCGTAATACACGATTTGTAATCGAATG 217
|||
Db 200 GGCTTTTAAACAAGCGGAAAAATACTAGCTAAAGATGTAATCATATCGAGCTAAT 259
|||
Qy 218 AAATGCAATTAATAAGTGGGTGCTGATTTAGAAAATGGTATTTTCTGCTGATT 277
|||
Db 260 TAATGAATGAATTAATAAATTCGATAAACAATAGCTCAAGGAATATATGATCGGATC 319
|||
Qy 278 ACGAATCCTTATATGATAATAGTACATATGCTTCTCACTTTTATGATCGGATATCG 337
|||
Db 320 ATAAAAATCCATATATGATAGTACTAGTACATTTTATCTCAATTTTATATCTGTAGAG 379
|||
Qy 338 GAACAACATATATCTTTTGGGAAACATGCAAAAGAAACAGCGCCAAATATTTTAACC 397
|||
Db 380 ATAATACCTTATTTGCGCGGTTTGTCTAATGCGAAAATTAACAGGACGAAGTATTTCAATC 439
|||
Qy 398 TTGCTGGTCAAGCATACCAAAATCAAGATATGCAAGAGCATTTCTTACTTAGGATTAT 457
|||
Db 440 AATCGTGACTGATTTACGAGAGGAAATTTGACACAGCGTTTATATAATTAGGCCTAG 499
|||
Qy 458 CGCTTCATTTATAGGAGATGTGAATCAGCAATGCAATGCAAGCAAACTTTTACGAATCTTT 517
|||
Db 500 CAATCCATTTATATACCGATATAGTCAACCTATGCAAGCAATATTTTACCGCAATAT 559
|||
Qy 518 CTTATCCAATGGGTTTCCATTTAAATACGAAAATTTTGTGATACATAAAAAATAACT 577
|||
Db 560 CATACCTCCAGGCTACCACCTGTGCATATGAAAATTTAGTAGATACCAATTAACACAAT 619
|||
Qy 578 ATATTGTTTCAGATAGCAATGATATTGGAATTTGAAAAGGACCAACCCAGAAGATTGGA 637
|||
Db 620 ATCAAGCAACGGAAGACATGGTAGCAAAAAGATTTTGCTCAGATGACGTGAAGACTGCG 679
|||
Qy 638 TTGAAGGAGCAGCGGTAGCAGCTAAACAAGATTTATCTGGCGTTTGTGAACGATAGCAAA 697
|||
Db 680 TCTATGAAAATGCGAAAAGGGGAAAGCGGACTACCCGAAAATAGTCAATGCGAAAACCTA 739
|||
Qy 698 AAGATTGGTTTGTAAAAGCA 717
|||
Db 740 AAAAATCATATTTTAGTAGGA 759
|||||
```

```
Db 740 AAAAATCATATTTTAGTAGGA 759
|||||
RESULT 13
AX415655
LOCUS AX415655 870 bp DNA linear PAT 02-SEP-2002
DEFINITION Sequence 2646 from Patent WO0228891.
ACCESSION AX415655
VERSION AX415655.1 GI:21448112
KEYWORDS
SOURCE Listeria monocytogenes EGD-e
ORGANISM Listeria monocytogenes EGD-e
REFERENCE 1
AUTHORS Kunst, F. and Glaser, P.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 2646 11-APR-2002;
INSTITUT PASTEUR (FR); CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES
source
Location/Qualifiers
1..870
/organism="Listeria monocytogenes EGD-e"
/mol_type="unassigned DNA"
/db_xref="taxon:169963"
ORIGIN
Query Match 18.2%; Score 154.4; DB 6; Length 870;
Best Local Similarity 53.1%; Pred. No. 8.2e-22;
Matches 329; Conservative 0; Mismatches 291; Indels 0; Gaps 0;
Qy 98 CACCGATTTTAAAGATGGTCAGCTGAGGATAAGCAATAATGAGGGGATTAACCTCTCATTTGT 157
|||
Db 140 CACATAAACTTAGTTGGTCCGGGATAACCGACAAATACTGACGTAAATAGCGACTATT 199
|||
Qy 158 GGATTGTAATCGTGCAATTGACATCATGCTCGTAATACACGATTTGTAATCGAATG 217
|||
Db 200 GGCTTTTAAACAAGCGGAAAAATACTAGCTAAAGATGTAATCATATCGAGCTAAT 259
|||
Qy 218 AAATGCAATTAATAAGTGGGTGCTGATTTAGAAAATGGTATTTTCTGCTGATT 277
|||
Db 260 TAATGAATGAATTAATAAATTCGATAAACAATAGCTCAAGGAATATATGATCGGATC 319
|||
Qy 278 ACGAATCCTTATATGATAATAGTACATATGCTTCTCACTTTTATGATCGGATATCG 337
|||
Db 320 ATAAAAATCCATATATGATAGTACTAGTACATTTTATCTCAATTTTATATCTGTAGAG 379
|||
Qy 338 GAACAACATATATCTTTTGGGAAACATGCAAAAGAAACAGCGCCAAATATTTTAACC 397
|||
Db 380 ATAATACCTTATTTGCGCGGTTTGTCTAATGCGAAAATTAACAGGACGAAGTATTTCAATC 439
|||
Qy 398 TTGCTGGTCAAGCATACCAAAATCAAGATATGCAAGAGCATTTCTTACTTAGGATTAT 457
|||
Db 440 AATCGTGACTGATTTACGAGAGGAAATTTGACACAGCGTTTATATAATTAGGCCTAG 499
|||
Qy 458 CGCTTCATTTATAGGAGATGTGAATCAGCAATGCAATGCAAGCAAACTTTTACGAATCTTT 517
|||
Db 500 CAATCCATTTATATACCGATATAGTCAACCTATGCAAGCAATATTTTACCGCAATAT 559
|||
Qy 518 CTTATCCAATGGGTTTCCATTTAAATACGAAAATTTTGTGATACATAAAAAATAACT 577
|||
Db 560 CATACCTCCAGGCTACCACCTGTGCATATGAAAATTTAGTAGATACCAATTAACACAAT 619
|||
Qy 578 ATATTGTTTCAGATAGCAATGATATTGGAATTTGAAAAGGACCAACCCAGAAGATTGGA 637
|||
Db 620 ATCAAGCAACGGAAGACATGGTAGCAAAAAGATTTTGCTCAGATGACGTGAAGACTGCG 679
|||
Qy 638 TTGAAGGAGCAGCGGTAGCAGCTAAACAAGATTTATCTGGCGTTTGTGAACGATAGCAAA 697
|||
Db 680 TCTATGAAAATGCGAAAAGGGGAAAGCGGACTACCCGAAAATAGTCAATGCGAAAACCTA 739
|||
Qy 698 AAGATTGGTTTGTAAAAGCA 717
|||
Db 740 AAAAATCATATTTTAGTAGGA 759
|||||
```

```
RESULT 14
LOCUS LISACTLDH
DEFINITION Listeria monocytogenes lecithinase, surface antigen. BCT 26-APR-1993
           gene complete cds, (plcB) gene complete cds, (ldh) gene complete
           cds.
ACCESSION M82881
VERSION 1
KEYWORDS GI:149644
SOURCE lactate dehydrogenase; lecithinase; surface antigen.
ORGANISM Listeria monocytogenes
REFERENCE Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
AUTHORS Vasquez-Boland J.A., Kocks, C., Dramsi, S., Ohayon, H., Geoffroy, C.,
Maugaud, J. and Cossart, P.
TITLE Nucleotide sequence of the lecithinase operon of Listeria
monocytogenes and possible role of lecithinase in cell-to-cell
spread
JOURNAL Infect. Immun. 60 (1), 219-230 (1992)
MEDLINE 92104678
PUBMED 1309513
COMMENT Original source text: Listeria monocytogenes (strain L028) DNA.
FEATURES
     source      Location/Qualifiers
     1..5648
         /organism="Listeria monocytogenes"
         /mol_type="genomic DNA"
         /strain="L028"
         /db_xref="taxon:1639"
     228..2147
         /gene="actA"
     228..2147
         /gene="actA"
         /codon_start=1
         /transl_table=1
         /protein_id="AAA25269.1"
         /db_xref="GI:149645"
         /translation="MGLNRFMRMMVVFITANCITINPDIIFAATSDSSSLNTDEWE
         EKTEQEVNTGPRYETAREVSRDIEKEFKSNKVRNTNKADLIAMLEKAKSGPN
         INNNSQETENAAINEEASGADPAIQVERHPGLPDSAAIEKKRKAIASSSELE
         SITYDPKPTKVKKVAKESVADASESDLSMQSADSSPOPLKANOQPPFPKVKK
         INDAGWVRDKDENPEVKAIVDKSAGLIDQLTKKSEVNASDFPPPTDEBELRL
         ALPETMLLGNAPATSPPSEFPPTDEBELRLALPETMLLGNAPATSPPSE
         FPPTELEIIRITASLDSLSPFRGDLALRNHNSQNSDFPPIPTPEELNGR
         GSRPSSEFSSNGDFTDENSETTEIEIDLADLRDRGTGKSRNAGFLPLNPFAS
         SPVSLSPVKISAPALISDITKTPFPNPSQPLNPNKTTTKTVTKPTPVKTPAP
         KLAELPATKQPTVLRNKTPPIEQAKTNKSINMPSLPVIOKEATESKEEMKQPT
         EKMVVESESANNANGKRSAGIEBGLIAKSAEDERAKKEPFGNHTLLILAMLAIGVF
         SLGAFIKIQLRNKN"
     2183..3052
         /gene="plcB"
     2183..3052
         /gene="plcB"
         /codon_start=1
         /transl_table=1
         /protein_id="AAA25270.1"
         /db_xref="GI:149646"
         /translation="MKFAKVVLMGLCLIASVLVFPVTIKANACDVEYLQTPAAPHDIDS
         KLPKLSWSANPNTDNTVHYMLFQKAEKILAKDVNHNMBANLNLKLPDKQIAQGI
         YDADHKNPYDTSTFLSHFYNPDRNTYLPGFANAKITGAKYFNQSVTDYREKGFDTA
         FYKGLAIHYTDTISQPMHANNFTAISYPGHCAYENYVDITKHNQATQEDMVAKRF
         CSDVDKWLYLENAKRAKADVPKIVNANTKSYLVGNSEWKDITVEPTGARLRDSQQL
         AGFLFLEFSKKTNE"
     3124..3540
         /misc_feature
     3582..4043
         /note="ORF/Y"
     3582..4043
         /note="ORFZ"
         /codon_start=1
         /transl_table=1
         /protein_id="AAA25271.1"
         /db_xref="GI:149647"
         /translation="MMMLKIGTAVLFVMIMAVMLVACGDKEETKTYTLTSQNGVDSKLT
         YTYKGDVTKTQATNTMSYASIGVASKEDAEMKMKATSDKPGQIDGLKEKLEYKDDKA
         IETLEVDYTKISSEDMNKIPGMSSNGDTSKGISMEESAOMLESQSIKEVSK"
terminator
CDS
     complement(4094..4426)
     /note="ORFB"
     /codon_start=1
     /transl_table=1
     /protein_id="AAA25272.1"
     /db_xref="GI:149648"
     /translation="MIVTSPNIGKQIIIEYKIVFGIEVITGVNFMKDIGAGLRNFFG
     GRSQYDELINAREEARIREMQAKDIGANAVIGVDIDYEVLGADNMLMVTASGTA
     WVIEAQDY"
     complement(4493..5167)
     /note="ORPA"
     /codon_start=1
     /transl_table=1
     /protein_id="AAA25273.1"
     /db_xref="GI:149649"
     /translation="MITFDQLDTLEQALENPNTIKIPRNHGPCDSDLDYGLKIGDLKK
     IIRREKLTQNHSLAVLIESNNSDLIYGLLAINPNKITTEQIEKNIAIFRETWSOLT
     FGLASIVSKRDDALLFAKTWIESDYDLTKSMHQIYSEHINNLEPBEATLLQRAKETIQ
     TETNRTYSMMNGPIITCGIYKDDLHEKAMEAKSVGKVNHLNLTACKVDFDAISYIEK
     ARNRTK"
     complement(5244..5648)
     /gene="ldh"
     /codon_start=1
     /transl_table=1
     /protein_id="AAA25274.1"
     /db_xref="GI:149650"
     /translation="EPNWSHTTVGSLPITEHISEDEQGMDFIFVSVRDAAYEIIINK
     KGATFYGVAAALARIKAILNENNAIPLSVLDLGHYGMNDIYIGAPAVVNRQGVRII
     VEHMLNDKEKEQMKNSADTLKVLDDAMKQID"
     complement(5247..5648)
     /gene="ldh"
Query Match      18.28; Score 154.4; DB 1; Length 5648;
Best Local Similarity 53.1%; Pred. No. 6.8e-22;
Matches 329; Conservative 0; Mismatches 291; Indels 0; Gaps 0;
Qy 98 CACCGATTTTAAAGATGTCAGCTGAGGATAAGCATATGAGGGGATTAACCTCTCATTTCT 157
Db 2322 CACATAAACTTAGTTGTCGCGGATAACCCGACAATAGTACGTAATAAGCCTATT 2381
Qy 158 GGATTGTAATCGTCAATGACATGTCGTGTAATCAACAGATGFGAAATCCGAATG 217
Db 2382 GGCTTTTAAACAAGCGGAAAAATACTAGCTAAAGATGATAATCATATCGAGCTAAT 2441
Qy 218 AAATCGCATTTATTAATGAGTGGCGTGTGATTAGAAAATCGTATTTATTTCTGCTGATT 277
Db 2442 TAATGAATGAATTAATAAATTCGATAACAATAGCTCAAGAAATATATGATCGGATC 2501
Qy 278 ACAGAAATCCTTATTATGATAATAGTACATATGCTTCTCAGCTTTTATGATCCGATAC 337
Db 2502 ATAAAAATCCATATTATGATAGTACTAGTACATTTTATCTCTATTTTATATCTCGATAG 2561
Qy 338 GAACAACATATATCTCTTTGCGAACAATGCAAGAAAGAACAGCCGCAAAATATTTTAACC 397
Db 2562 ATAAACTATTATTGCGGGTTTGTCTAAATGCGAAATTAACAGGAGCAAGTATTTTCAATC 2621
Qy 398 TTGCTGGTCAACATACCAAAATCAAGATATGACAGAACGATCTTCTTACTTAGGATTAT 457
Db 2622 AATCGGTGACTGATTACCGAGAAAGGAAATTTGACACAGCGTTTTTATAAATAGGCCTAG 2681
Qy 458 CGCTTCATTATTAGGAGATGTAATCAGCCATGCAATGATGAGCAAACTTTAGGAATCTTT 517
Db 2682 CAATCCATTATTATACGGAATATTAGTCAACCTATGACCGCAATAATTTTACCGCAAT 2741
Qy 518 CTTATCCAATGGTTTCCATTCTTAATACGAAAAATTTTGTGTATACAAATAAAAAATAACT 577
Db 2742 CATACCCCTCCAGGCTACCCTGTGCATATGAAAAATACGTAGATACCAATTAAACACAT 2801
Qy 578 ATATTGTTTCAGATAGCAATCGATATTGGAAATGGAAGGAGCAAAACCCAGAAATTTGGA 637
```

Db 2802 ATCAAGCAACGGAAGACATGTGTAGCAAAAGATTTTGTCTCAGATGACGTGAAGACTGGC 2861

Qy 638 TTGAAGGAGCGCGTAGCAGCTAAACAGATTATCTCTGGCTTGTGAACGATACGACAA 697

Db 2862 TCTATGAAATGCGAAGGCGAAGCGGACTACCGGAAATATCTCAATGCGAAACTA 2921

Qy 698 AAGATTGGTTGTTAAAGCA 717

Db 2922 AAAATCATATTTTAGTAGGA 2941

RESULT 15

AL591974

LOCUS

DEFINITION

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

AL591974

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2004, 02:14:50 ; Search time 483 Seconds

(without alignments)
8618.563 Million cell updates/sec

Title: PCT-US03-12556-1

Perfect score: 849
Sequence: 1 atgaaaagaagattagc.....atagctatgtaaatgctaa 849

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3222919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	849	100.0	849	16	US-10-421-654-1
2	559.2	65.9	1192	16	US-10-369-493-40434
3	529	62.3	852	16	US-10-421-654-3
4	505.8	59.6	843	16	US-10-421-654-5
5	154.4	18.2	870	16	US-10-398-221-740
6	154.4	18.2	870	16	US-10-398-221-2646
7	87.4	10.3	1422	16	US-10-421-654-31
8	85.8	10.1	1422	16	US-10-421-654-97
9	76.8	9.0	1422	16	US-10-421-654-89
10	72	8.5	1422	16	US-10-421-654-81
11	57.4	6.8	4562	16	US-10-398-221-3772
12	43.4	5.1	3821	8	US-08-781-986A-470
13	43.4	5.1	9821	13	US-10-329-624-470
14	43.2	5.1	837	10	US-09-966-521-21

c 15	43.2	5.1	837	15	US-10-429-094-21	Sequence 21, Appl
c 16	43.2	5.1	5924	8	US-08-781-986A-130	Sequence 130, App
c 17	43.2	5.1	5924	13	US-10-329-624-130	Sequence 130, App
c 18	43	5.1	8170	15	US-10-240-453-132	Sequence 132, App
c 19	41.8	4.9	2000	16	US-10-260-238-1941	Sequence 1941, App
c 20	41.8	4.9	6227	15	US-10-311-455-1559	Sequence 1559, App
c 21	41.8	4.9	6227	17	US-10-240-589C-69	Sequence 69, Appl
c 22	41.6	4.9	2319	14	US-10-067-385-7	Sequence 7, Appl
c 23	41.6	4.9	2911	13	US-10-158-844-199	Sequence 199, App
c 24	41.6	4.9	6360	10	US-09-769-744A-27	Sequence 27, Appl
c 25	41.6	4.9	6423	13	US-10-282-122A-37486	Sequence 37486, A
c 26	41.6	4.9	495269	16	US-10-398-221-8	Sequence 8, Appl
c 27	41.6	4.9	3011208	16	US-10-398-221-2058	Sequence 2058, Ap
c 28	41.2	4.9	3673778	15	US-10-312-841-2	Sequence 2, Appl
c 29	41	4.8	4985	15	US-10-094-240-10	Sequence 10, Appl
c 30	41	4.8	4985	15	US-10-056-405-10	Sequence 10, Appl
c 31	40.4	4.8	541	13	US-10-424-599-52877	Sequence 52877, A
c 32	40.2	4.7	1014	13	US-10-282-122A-31009	Sequence 31009, A
c 33	40	4.7	3381	9	US-09-815-242-4318	Sequence 4318, Ap
c 34	40	4.7	3477	9	US-09-815-242-8424	Sequence 8424, Ap
c 35	39.8	4.7	1353	15	US-10-349-680-131	Sequence 131, App
c 36	39.8	4.7	2130	13	US-10-282-122A-40566	Sequence 40566, A
c 37	39.6	4.7	3673778	15	US-10-312-841-1	Sequence 1, Appl
c 38	39.4	4.6	5338	13	US-10-221-613-241	Sequence 241, Appl
c 39	39.2	4.6	13038	15	US-10-311-455-1247	Sequence 1247, Ap
c 40	39.2	4.6	60461	16	US-10-341-434-82	Sequence 82, Appl
c 41	39.2	4.6	150351	17	US-10-322-281-453	Sequence 453, App
c 42	39	4.6	835	15	US-10-029-386-22786	Sequence 22786, A
c 43	39	4.6	16217	15	US-10-311-455-597	Sequence 597, App
c 44	39	4.6	640681	9	US-09-790-988-1	Sequence 1, Appl
c 45	38.8	4.6	2583	17	US-10-441-926-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-10-421-654-1
; Sequence 1, Application US/10421654
; Publication No. US20040005604A1
; GENERAL INFORMATION:
; APPLICANT: Gramatikova, Svetlana
; APPLICANT: Hazlewood, Geoff
; APPLICANT: Lam, David E.
; APPLICANT: Barton, Nelson R.
; TITLE OF INVENTION: PHOSPHOLIPASES, NUCLEIC ACIDS ENCODING THEM AND
; TITLE OF INVENTION: METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-094001
; CURRENT APPLICATION NUMBER: US/10/421,654
; CURRENT FILING DATE: 2003-04-21
; PRIOR APPLICATION NUMBER: US 60/374,313
; PRIOR FILING DATE: 2002-04-19
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 849
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-421-654-1

Query Match 100.0%; Score 849; DB 16; Length 849;
Best Local Similarity 100.0%; Pred. No. 5.6e-196;
Matches 849; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAAAAAGAAAGTATTAGCACTAGCAGCTATGTTGCTTTAGTGCACCAAGTTCACAGT 60
Db 1 ATGAAAAAGAAAGTATTAGCACTAGCAGCTATGTTGCTTTAGTGCACCAAGTTCACAGT 60
QY 61 GTAGTATTTCACAAACAATAATAGTCAAGTCTGCACCAAGTTCAGTTCAGTTCAGT 120
Db 61 GTAGTATTTCACAAACAATAATAGTCAAGTCTGCACCAAGTTCAGTTCAGTTCAGT 120

QY 121 GAGGATAGCATATAGGGGATTAACCTCTGTTGGAATGTAATAATCGTGCAATTGAC 180
Db 121 GAGGATAGCATATAGGGGATTAACCTCTGTTGGAATGTAATAATCGTGCAATTGAC 180
QY 181 ATCATGCTCTGTAATACCAAGATGTTGAATCGGAATGAATATGCAATTAATAATGAGTGG 240
Db 181 ATCATGCTCTGTAATACCAAGATGTTGAATCGGAATGAATATGCAATTAATAATGAGTGG 240
QY 241 CGTCTGATTTAGAAATGTTATTTCTGCTGATTTACGGAATCCCTTATTAATGATAAT 300
Db 241 CGTCTGATTTAGAAATGTTATTTCTGCTGATTTACGGAATCCCTTATTAATGATAAT 300
QY 301 AGTACATATGCTCTTACCTTTATGATCCGGATACCTGGAACAAATATATTTCCCTTTTGGC 360
Db 301 AGTACATATGCTCTTACCTTTATGATCCGGATACCTGGAACAAATATATTTCCCTTTTGGC 360
QY 361 AAACATGCAAAAGAAACAGGGCGCAAAATATTTTAACTTGTGTCGATCAAGCATACCAAAAT 420
Db 361 AAACATGCAAAAGAAACAGGGCGCAAAATATTTTAACTTGTGTCGATCAAGCATACCAAAAT 420
QY 421 CAAGATATGCAAGCAATCTTCTACTAGGATTTATCGCTTATCGCTTATTTAGGAGATGTG 480
Db 421 CAAGATATGCAAGCAATCTTCTACTAGGATTTATCGCTTATTTAGGAGATGTG 480
QY 481 AATCAGCAATGCAATGCAAGCAATCTTCTACTAGGATTTATCGCTTATTTAGGAGATGTG 540
Db 481 AATCAGCAATGCAATGCAAGCAATCTTCTACTAGGATTTATCGCTTATTTAGGAGATGTG 540
QY 541 AAATACGAAATTTTGTGATACAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 600
Db 541 AAATACGAAATTTTGTGATACAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 600
QY 601 TATTGGATTTGGAAGGAGCAACCCAGAGATTTGATTTGGAAGGAGCAACCCAGAGATTTG 660
Db 601 TATTGGATTTGGAAGGAGCAACCCAGAGATTTGATTTGGAAGGAGCAACCCAGAGATTTG 660
QY 661 AAACAGGATTTCTGCGGTTGTGAACGATACGACAAAGATTTGTTTGTAAAGCAGCC 720
Db 661 AAACAGGATTTCTGCGGTTGTGAACGATACGACAAAGATTTGTTTGTAAAGCAGCC 720
QY 721 GTATCTCAAGATATGCAAGATTTGGAACGATACGACAAAGATTTGTTTGTAAAGCAGCC 780
Db 721 GTATCTCAAGATATGCAAGATTTGGAACGATACGACAAAGATTTGTTTGTAAAGCAGCC 780
QY 781 TTAATGGAAGCGCAGCGGTTACAGCTGTTATATTTTATTTTATTTTATTTTATTTTATTT 840
Db 781 TTAATGGAAGCGCAGCGGTTACAGCTGTTATATTTTATTTTATTTTATTTTATTTTATTT 840
QY 841 AATCGCTAA 849
Db 841 AATCGCTAA 849

RESULT 2

US-10-369-493-40434
; Sequence 40434, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 40434

; LENGTH: 1192

; TYPE: DNA

; ORGANISM: Bacillus thuringiensis

US-10-369-493-40434

Query Match 65.9%; Score 559.2; DB 16; Length 1192;

Best Local Similarity 79.3%; Pred. No. 1.7e-125;

Matches 676; Conservative 0; Mismatches 173; Indels 3; Gaps 1;

QY 1 ATGAAAAGAAAGATTTAGCACTAGCAGCTATGCTTGTAGCTGCGCCAGTTCAAGT 60
Db 1 ATGAAAAGAAAGATTTAGCACTAGCAGCTATGCTTGTAGCTGCGCCAGTTCAAGT 60
QY 61 GTAGTATTTGCAAAACAAATAAATAGT---AAAAGTCTCTGCACCGATTTTAAAGATGTGCA 117
Db 61 GTAGTATTTGCAAAACAAATAAATAGT---AAAAGTCTCTGCACCGATTTTAAAGATGTGCA 117
QY 118 GCTGAGATATAGCATATAGGGGATTAACCTCTCACTTTGTGATTTAAATCGTGCAATT 177
Db 121 GCTGAGATATTAACATAAAGAGGCGTGAATTTCTCATTTATGGAATTGTAAACCGTGCAATT 180
QY 178 GACATCATGCTCTGTAATACAAACGATTTGTAATCGGAATGAACTGCATTTATAAATGAG 237
Db 181 GATATTATGCTCTGTAATACAAACGATTTGTAATCGGAATGAACTGCATTTATAAATGAG 240
QY 238 TGGCGTCTGATTTAGAAAATGGTATTTATTTCTGCTGATTAACGAAATCCCTTATATGAT 297
Db 241 TGGCGTCTGATTTAGAAAATGGTATTTATTTGCTGCTGATTAACGAAATCCCTTATATGAT 300
QY 298 AATAGTACATATGCTTCTCACTTTTATGATCCGGATCTGGAACAAATATATTTCCCTTTT 357
Db 301 AATAGTACATATGCTTCTCACTTTTATGATCCGGATCTGGAACAAATATATTTCCCTTTT 360
QY 358 GCGAAAACATGCAAAAGAAACAGGGCGCAAAATATTTTAACTTGTGCTGCTCAAGCATACCA 417
Db 361 GCAAGCAGGCAAAAGAAACCTGGAGCTAAATATTTTAAATAGTGTGATGCTTTACAA 420
QY 418 AATCAAGATATGCAAGCAATCTTCTTACTAGGATTTATCGCTTCAATTTATAGGAGAT 477
Db 421 AATCAAGATATGCAAGCAATCTTCTTACTAGGATTTATCGCTTCAATTTATAGGAGAT 480
QY 478 GTGAATCAGCAATGCAATGCAAGCAATCTTCTTACTAGGATTTATCGCTTCAATTTCCAT 537
Db 481 GTGAATCAGCAATGCAATGCAAGCAATCTTCTTACTAGGATTTATCGCTTCAATTTCCAT 540
QY 538 TCTAAATACGAAAATTTTGTGATACAAATAAATAAATAAATAAATAAATAAATAAATAAATA 597
Db 541 TCTAAATACGAAAATTTTGTGATACAAATAAATAAATAAATAAATAAATAAATAAATAAATA 600
QY 598 GGATATTTGGAATTTGGAAGGAGCAACCCAGAGATTTGGAATTTGGAAGGAGGAGCGGTAGCA 657
Db 601 GGATATTTGGAATTTGGAAGGAGGTAACAAATCCAGAGATTTGGAATTTGGAAGGAGGAGTGT 660
QY 658 GCTAAAACAGATTTATCTGCGGTTGTGAACGATACGACAAAGATTTGTTTGTAAAGCA 717
Db 661 GCGAAAACAGATTTATCGTGGCATTTGTAATGATTAATACGAAAGATTTGGTTGTAAGAGCA 720
QY 718 GCGGTATCTCAAGATATGCAAGATTAATGCGGCTGCGGAAGTAAACACCGGTGACAGGAAAG 777
Db 721 GCTGTATCAAGATATGCAAGATTAATGCGGCTGCGGAAGTAAACACCGGTGACAGGTAAG 780
QY 778 CGTTTAAATGGAAGCGCAGCGGTTTACAGCTGTTTATATTTTCAATTTTGTGTTGATACGAT 837
Db 781 CGTTTAAATGGAAGCGCAGCGGTTTACAGCTGTTTATATTTTCAATTTTGTGTTGATACGAT 840
QY 838 GTAATCGCTAA 849
Db 841 GGAATCGCTAA 852

RESULT 3

US-10-421-654-3
; Sequence 3, Application US/10421654

```
; Publication No. US20040005604A1
; GENERAL INFORMATION:
; APPLICANT: Gramatikova, Svetlana
; APPLICANT: Hazlewood, Geoff
; APPLICANT: Lam, David E.
; APPLICANT: Barton, Nelson R.
; TITLE OF INVENTION: PHOSPHOLIPASES, NUCLEIC ACIDS ENCODING THEM AND
; FILE REFERENCE: 09010-094001
; CURRENT APPLICATION NUMBER: US/10/421,654
; CURRENT FILING DATE: 2003-04-21
; PRIOR APPLICATION NUMBER: US 60/374,313
; PRIOR FILING DATE: 2002-04-19
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 852
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; US-10-421-654-3

Query Match      62.3%; Score 529; DB 16; Length 852;
Best Local Similarity 76.4%; Pred. No. 3.2e-118; Indels 0; Gaps 0;
Matches 649; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

Qy 1 ATGAAAAAGAAAGTATTAGCACTAGCAGCTATGGTTGCTTTAGCTGCGCAGCTTCAAAGT 60
Db 1 ATGAAAAAGAAATTTAGCTATAGCTCCGTAATGCTTTTAACAGCTCTTATCCAAAGT 60

Qy 61 GTAGTATTGGCAAAACAAATAATAGTGAAGTCTCTGACCGATTTTAAAGTGGTCAGCT 120
Db 61 GTGGCGTTTGGCGCATGAAATGGTCAACCAAGTCAACCAATTTGCTTAAAGTGGTCAGCA 120

Qy 121 GAGGATAGCAATAATGAGGGGATTAACCTCATTTTGTGATTTGTAATCGTCAATTGAC 180
Db 121 GAATCTATACATAATGAAGGAGTAAAGTCTCATTTTATGATTTGTAACAGAGCCATTGAT 180

Qy 181 ATCATGCTCTCGTAATACAAACGATTTGTAATCGAATGAACTGCAATTTAAATGAGTGG 240
Db 181 ATTATGTCCTCAAAATACGCTGTTGTAAGCAAAATGACAGCTCTTAAATGAGTGG 240

Qy 241 CGTGTGATTAGAAAATGGTATTATTCTGCTGATTAGAGAAATCTTTATTTATGATAAT 300
Db 241 CGTACGGATCTAGAGAAAGGCAATTTACTCTGGCGATTATGAAACCCATCTATGATAAT 300

Qy 301 AGTACATATGCTTCTCATTATGATCCGGATCTGGAACCAACATATTTCTTTTGG 360
Db 301 TCCACATTTGCTTCACTTCTATGATCCTGATTCAGGAAACGTAATTTCCATTTGCT 360

Qy 361 AAACATGCAAAAGAAACAGGCGCAAAATATTTTAACTTGTCTGCTCAAGCATACCATAAT 420
Db 361 AAACAGCAAGCAAAACAGGAGCGAAATATTTTAAATGATGTTGAGCTTATCAAAT 420

Qy 421 CAAGATATGCAAGCAAGCTTCTTACTTATGAGTATTCGCTTCAATTTTAGGAGATGG 480
Db 421 AAAGATCTGAAACAGCAATCTTTTATTTAGGATTTATCACTTCACTATTTAGGGATGTC 480

Qy 481 AATGAGCAATGCTATGCAAGCAAACTTTTACGAATCTTCTTATTCGAATGGGTTTCACTT 540
Db 481 AACCAACCAATGCTATGCAAGCAAACTTTTACTAATATTTTGCATCTTCTTCACTCA 540

Qy 541 AAATACGAAAAATTTTGTGATACAAATAAAAAATACTATTTTGTTCAGATAGCAATGGA 600
Db 541 AAATATGAAAAATTTTGTGATACAGTGAAGCAATTTATAGTAAACGATGGAATGGC 600

Qy 601 TATTGGAATTTGAAAGGAGCAAAACCCAGAGATTTGGATTGAAAGGAGCAGCGTAGCAGCT 660
Db 601 TATTGGAATTTGCAAAAGTGAATTTCCAGAGAGTGGTTTCATGTCATCAGCATCAGCAGCA 660

Qy 661 AAACAGAAATTTCTGGGCTTTGTGAACGATACGACAAAGATTTGGTTTGTAAAGCAGCC 720
Db 661 AAACAGAAATTTCTGGGCTTTGTGAACGATTTTAAATTTAGCTTGTGAGAGCTTATCAAAT 720

Db 661 AAAGCTGATTTTCCATCAATTTGTTAATGATAGACGAAAAATTTGTTCTCTAAAGCAGCT 720
Qy 721 GTATCTCAAGATATGCAATAAATGCGGTGCGGAAGTAACACCGGTGACAGGAAAGCT 780
Db 721 GTATCAAGAAGTCTGCTGATATAAATGCGGTGCGGAAGTAACACCGATACAGGAAAAAGCT 780
Qy 781 TTAATGGAAGCGCAGCGGTTACAGCTGGTTATATTTCACTTTGTTGTTGATACGTTATGTA 840
Db 781 TTAATGGAAGCGCAGCGGTTTACAGCTGGATATATCCATTTATGGTTTGTGATACGTCGTCG 840

Qy 841 AATCGCTAA 849
Db 841 AATAACAAA 849

RESULT 4
US-10-421-654-5
; Sequence 5, Application US/10421654
; Publication No. US20040005604A1
; GENERAL INFORMATION:
; APPLICANT: Gramatikova, Svetlana
; APPLICANT: Hazlewood, Geoff
; APPLICANT: Lam, David E.
; APPLICANT: Barton, Nelson R.
; TITLE OF INVENTION: PHOSPHOLIPASES, NUCLEIC ACIDS ENCODING THEM AND
; FILE REFERENCE: 09010-094001
; CURRENT APPLICATION NUMBER: US/10/421,654
; CURRENT FILING DATE: 2003-04-21
; PRIOR APPLICATION NUMBER: US 60/374,313
; PRIOR FILING DATE: 2002-04-19
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 843
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; US-10-421-654-5

Query Match      59.6%; Score 505.8; DB 16; Length 843;
Best Local Similarity 75.7%; Pred. No. 1.4e-112;
Matches 643; Conservative 0; Mismatches 197; Indels 9; Gaps 1;

Qy 1 ATGAAAAAGAAAGTATTAGCACTAGCAGCTATGGTTGCTTTAGCTGCGCAGCTTCAAAGT 60
Db 1 ATGAAAAAGAAATTTTATGCTATAGCTTCTGTAATTTGCTTTAAAGCTCTTATTCNAAGT 60

Qy 61 GTAGTATTGGCAAAACAAATAATAGTGAAGTCTCTGACCGATTTTAAAGTGGTCAGCT 120
Db 61 GTGGCGTTTGGCGCATGAAATCTGATGGGCTATTGC-----TTTAAAGATGGTCAGCG 111

Qy 121 GAGGATAGCAATAATGAGGGGATTAACCTCATTTTGTGATTTGTAATCGTCAATTGAC 180
Db 112 GAATCTGATACATAATGAAGGAGTAAAGTTCTCATTTTGTGATTTGTAACAGAGCAATGAT 171

Qy 181 ATCATGCTCTCGTAATACAAACGATTTGTAATCGAATGAACTGCAATTTTAAATGAGTGG 240
Db 172 ATTATGTCCTCAAAATACGCTGTTGTAAGCAAAATGAGAGCTCTTATTAATGAGTGG 231

Qy 241 CGTGTGATTAGAAAATGGTATTTTATTTCTGCTGATTACGGAATCTTTATTTATGATAAT 300
Db 232 CGTACGAATTTGGAGGAAGGTATTTATTTCTGAGATTTATAAAACCCATCTATGATAAT 291

Qy 301 AGTACATATGCTTCTCATTCTTTATGATCCGGATCTGGAACACATATATTCCTTTTGG 360
Db 292 TCCACATTTGCTTCACTTCTATGATCTCTGATTCAGAAAAAACGATATATTTCCATTTGCT 351

Qy 361 AAACATGCAAAAGAAACAGGCGCAAAATATTTTAACTTTGCTGTCAGAGCATACCAAAAT 420
Db 362 AAACAGCAAGCAAAACAGGAGCAAGTATTTTAAATTTAGCTTGTGAGAGCTTATCAAAT 411
```

421	QAAGATATGCAGCAAGCAATTCCTCTA	CTTAGGATTATCGCTTCA	TATTTAGGAGATGTG	480
412	AAAGATCTGAAAAATGCAATTCCTTTAT	TTAGGATTA	TCCTTCA	471
481	AATCAGCCAAATGCATGCAGCAAACTTT	TACGAATCTTTCTTAT	CCAAATGGGTTC	540
472	AATCAACCAATGCATGCAGCAAACTTT	TAACATTTGCATCCA	TTCGCTTCCA	531
541	AAATAGCAAAATTTTGTGTGATACATA	AAAAAATAACTATATAT	TGTTTCAGATAGCAATGGA	600
532	AAATATGAAAACTTCGTGTGATACAGT	GAAAGACAATATATAGAGTAA	CAGATGGAGATGGC	591
601	TATTTGGAATTTGGAAGGAGCAAA	CCACAGACATTTGGAAT	TGAAGGAGCAGCGTAGCAGCT	660
592	TATTTGGAATTTGGAANAAGTGCAAT	CCAGAGAGTGGGTTCAT	GTGATCAGCATCAGCAGCA	651
661	AAACAGATTAATCTCGGGGTTGTG	AACATACGACAAAGAAT	TGGTTTGTAAAAGCAGCC	720
652	AAAGCTGAATTTCCCATCAATTTGTT	AATGATAATACGAAAAGTT	TGGTTTCTTAAAAGCAGCG	711
721	GTATCTCAAGAAATATGCAGATAAAT	TGGCGTCCGGAAGTAA	CACCGGTGACAGAAACGCT	780
712	GTATCACAAGACTCTGCTGACAAAT	TGGCGTCTGAAGTAA	CACCGGTAAACAGGAAAACGCT	771
781	TTAATCGAAGCGCAGCGGTTACAGCT	GGTTATATTCATTTGTGGTTTGAT	TAGTATGTA	840
772	TTAATCGAAGCAGCGGATTAACAGCT	TGATATATTCATTTATGGTTTGAT	ATACGTACGCTG	831
841	AATCGCTAA	849		
832	AATAACAAA	840		

```

RESULT 5
US-10-398-221-740
, Sequence 740, Application US/10398221
, Publication No. US20040018514A1
, GENERAL INFORMATION:
, APPLICANT: KUNST, Frederik
, APPLICANT: GLASER, Philippe
, TITLE OF INVENTION: Listeria innocua, genome and applications
, FILE REFERENCE: 344 702 - US
, CURRENT APPLICATION NUMBER: US/10/398,221
, CURRENT FILING DATE: 2003-03-27
, PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
, PRIOR FILING DATE: 2001-10-04
, PRIOR APPLICATION NUMBER: FR 00/12 697
, PRIOR FILING DATE: 2000-10-04
, NUMBER OF SEQ ID NOS: 4025
, SOFTWARE: PatentIn version 3.0
, SEQ ID NO 740
, LENGTH: 870
, TYPE: DNA
, ORGANISM: Listeria monocytogenes-EGD
US-10-398-221-740

```

	Query Match	18.2%	Score 154.4	DB 16	Length 870	
	Best Local Similarity	53.1%	Pred. No. 3.3e-27			
	Matches 329	Conservative	0	Mismatches 291	Indels	Gaps 0
QY	98	CACCGATTTTAAGATGGTCAGCTGAGGATAAGCATAAATGAGGGGATTAACCTCTCATTTGT	157			
Db	140	CACATAAACTTAGTTGGTCCGCGGATAACCGACAATACTAGCGTAATAATCGCACTATT	199			
QY	158	GGATTGTAATTCGTGCAATTGACATCATGTCTCGTAATACAACAGATTGTGAATCCGAATG	217			
Db	200	GGCTTTTTTAAACAACGCGGAAAAATACTAGCTAAAGATGTAAATCATATCGAGCTAAAT	259			
QY	218	AAACTGCAATTATAATGAGTGCGCTGCTGTTATAGAAAAATGGTATTTATTCCTGCTGATT	277			
Db	260	TAATGAATGAACCTTAAAAAATTCGATAAACAAATAGCTCAAGGAATATATGATCGGGATC	319			
QY	278	ACGAGAATCCCTTATTATGATTAATAGTACATATGCTTCTTCACCTTTTATGATCCGGATACGTG	337			

Db 320 ATAAAAATCCATATTATGATACCTAGTACATTTTATCTCATTTTATTAATCCTGATAGAG 379

Qy 338 GAACAAACATATATTCCTTTTTCGCAAAACATGCAAAAGAAACAGCGCGCAAAATATTTTAACC 397

Db 380 ATAATACTTATTTGCCGGGTTTTCGTAATCGGAAAAAACAAGGAGCAAAAGTATTTCAATC 439

Qy 398 TTGCTGTGAAGCATATACAAATATCGAATATGCGAAGCAATTCCTCTACTTAGGATTAT 457

Db 440 AATCGGTGACTGATTACCGAAGAGGGAATTTTGCACAGCGTTTATATAAATAGGCGCTAG 499

Qy 458 CGCTTCATATTTAGGAGTGTGAATCAGCCAATGCATGCAGCAAACTTTTACGAATCTTT 517

Db 500 CAATCCCATTTATACGGATATTAGTCAACCTATGCGACGCCAAATTTTACCGCAATAT 559

Qy 518 CTTATCCAAATGGGTTTCCTATCTAAATACGAAAAATTTTGTGTGATACATAAATAAATAACT 577

Db 560 CATACCTCCAGGCTACCATCTGTGCATATGAAAAATTACGTAGATACCATTAACACATTT 619

Qy 578 ATATTGTTTCAGATAGCAATGGATATTGGAAATTCGAAAGGAGCAAAACCCAGAAGATTGGA 637

Db 620 ATCAAGCAACGGAAGACATGGTACCAAAAAGATTTTGTCTCAGATGACGTGAAAGACTGGC 679

Qy 638 TTGAAGGAGCGCGGTAGCAGCTAAACAAGATTATCTTGGCGTTGTGAAACGATACGCAA 697

Db 680 TCTATGAAATGCGAAAGGCGGAAGCGGACTACCCGAAATATAGTCAATCGGAAACTA 739

Qy 698 AAGATTGGTTGTAAAAAGCA 717

Db 740 AAAAATCATATTAGTAGGA 759

RESULT 6

US-10-398-221-2646

; Sequence 2646, Application US/10398221

; Publication No. US20040018514A1

; GENERAL INFORMATION:

; APPLICANT: KUNST, Frederik

; APPLICANT: GLASER, Philippe

; TITLE OF INVENTION: *Listeria innocua*, genome and applications

; FILE REFERENCE: 344 702 - US

; CURRENT APPLICATION NUMBER: US/10/398,221

; CURRENT FILING DATE: 2003-03-27

; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061

; PRIOR FILING DATE: 2001-10-04

; PRIOR APPLICATION NUMBER: FR 00/12 697

; PRIOR FILING DATE: 2000-10-04

; NUMBER OF SEQ ID NOS: 4025

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 2646

; LENGTH: 870

; TYPE: DNA

; ORGANISM: *Listeria monocytogenes* EGDe

US-10-398-221-2646

```

RESULT 6
US-10-398-221-2646
; Sequence 2646, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2646
; LENGTH: 870
; TYPE: DNA
; ORGANISM: Listeria monocytogenes EGDe
US-10-398-221-2646

```

	Query Match	18.2t; Score 154.4; DB 16; Length 870;
	Best Local Similarity 53.1t; Pred. No. 3.3e-27;	
	Matches 329; Conservative 0; Mismatches 291; Indels 0; Gaps 0;	
QY	98 CACCGATTTTAAAGATGGTTCAGCTCAGAGATAAGCATATAATGAGGGGATTAACCTCTCATTTGT 151	
DB	140 CACATAAATTTAGTTGGTCCCGGATTAACCGACAAATACTGACGTAAATACGCACATAAT 199	
QY	158 GGATTGTAAATCGTGCATTTGACATCATGTCTCGTAATACAAAGATTGTGAATCCGGAATG 217	
DB	200 GCTTTTAAACAACGCGGAAAAATACTAGCTAAAGATGTAATCATATCGGAGCTAAAT 259	
QY	218 AAACCTGCATTTTAATAGTGGCGTCTGATTTTAGAAAATGGTATTTATCTCGTCGATT 277	
DB	260 TAATGAATGAACTTTAAAAATTCGATTAACAACAATAGCTCAGGGAATATATGATCCGGATC 319	
QY	278 ACGAGAATCCCTATTATGATAAATAGTACATATGCTTCTCACTTTTATGATCCGGATATCTG 337	

Db 320 ATAAAAATCCATATATATGATAGTACTAGTACATTTTATCTCATTTTATATCTCTGATAGAG 379
Qy 338 GAACACATATATCTTTTGGAAACATGCAAAAGAACAGCGCCGCAAAATATTTTAACC 397
Db 380 ATAATACATTTATTTGGCGGTTTGTCTAATGCGAAATAACAGGAGCAAGATATTTCAATC 439
Qy 398 TTGCTGTCACAGCATACCAAAATCAAGATATGCAAGCAAGCATCTTCTACTTAGGATAT 457
Db 440 AATCGGTGACTGATTTACCGAGAAGGGAATTTGACACAGCGTTTATTAATTTAGCCCTAG 499
Qy 458 CGCTTCATTTATTTAGGAGATGTAATCAGCCCAATGATGCAAGCAAACTTTTACGAATCTTT 517
Db 500 CAATCCATTTATATACGATATTTAGTCAACCTATGCAAGCCCAATATTTTACCGCAATAT 559
Qy 518 CTTATCCAAATGGTTTCCATCTAAATACGAAATTTTGTGATACAATAAAAAATAACT 577
Db 560 CATACCTCCAGGCTACCACTGTGATATGAAATTTAGCTAGTACCACTTTAAACACAAT 619
Qy 578 ATATTGTTTCAGATAGCATGATATTTGGAATTTGAAAGGAGCAACCCAGAGATTGA 637
Db 620 ATCAAGCAACGGAAGACATGTTAGCAAAAGATTTTGTCTCAGATGACGTTGAAAGACTGCG 679
Qy 638 TTGAAGGAGCAGCGGTAGCAGCTAAACAAAGATTTATCTTGGCGTTGTGAACGATACGACAA 697
Db 680 TCTATGAAATGCGAAAGGCGAAAGCGGACTACCCGAAATAGTCAATGCGAAACTA 739
Qy 698 AAGATTGGTTTGTAAAGCA 717
Db 740 AAAATCATATTTAGTAGGA 759

RESULT 7

US-10-421-654-31
; Sequence 31, Application US/10421654
; Publication No. US20040005604A1
; GENERAL INFORMATION:
; APPLICANT: Gramatikova, Svetlana
; APPLICANT: Hazlewood, Geoff
; APPLICANT: Lam, David E.
; APPLICANT: Barton, Nelson R.
; TITLE OF INVENTION: PHOSPHOLIPASES, NUCLEIC ACIDS ENCODING THEM AND
; TITLE OF INVENTION: METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-094001
; CURRENT APPLICATION NUMBER: US/10/421,654
; CURRENT FILING DATE: 2003-04-21
; PRIOR APPLICATION NUMBER: US 60/374,313
; PRIOR FILING DATE: 2002-04-19
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 1422
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-421-654-31

Query Match 10.3%; Score 87.4; DB 16; Length 1422;
Best Local Similarity 53.3%; Pred. No. 8.2e-11;
Matches 284; Conservative 0; Mismatches 231; Indels 18; Gaps 4;
Qy 313 TCTCACTTTTATGATCCGGATACCTGGAAACAACATATATTTCCCTTTTTCGAAACATGCAAAA 372
Db 385 TCTCATTTCTATGATCCTGATAGCAAGAAAGAACTATAAGGAGAGAGAGAACCAACAGCC 444
Qy 373 GAAAC---AGCGGCAAAATATTTTAACTTGTCTGCTCAAGCATACCAAAATCAAGATATG 429
Db 445 CTTTCCGAGGGGATAAATATTTTAAATTTAGCAGGAGATTTATTTAAGAAAGAAAGATTGG 504
Qy 430 CAGCAAGCATTTCTTCTACTAGGATTTATCGCTTCAATTTTATAGGAGATGTTGAATCAGCCA 489
Db 505 AAACAAGCTTTCTATTTATTTAGGTGTCGACGCACTTACTTACAGATGCTACTCAGCCA 564

Qy 490 ATGCATGCGAGCAAACTTTTACGAATCT---TTCTTATCCAAATGGGTTTTCATTTCTAAATAC 546
Db 565 ATGCATGCTGCTAAATTTTACAGCTGTGACATGAGTGCATTAAGCTTTTTCATAGCGCTTTT 624
Qy 547 GAAATTTTGTGATACAAATAAATAAATTAATATATTTTGTTCAGATAGCAATCGATATTGG 606
Db 625 GAAATTTATGTAACGACAGTTTCAGACACCGTTTGAAGTGAAGGATGATAGGGAACATAT 684
Qy 607 AATTGGAAGGAGCAAAACCCAGAGATTGGATTTGAAGGAGCAGCGGTAGCAGCTAAACAA 666
Db 685 AATTGG-----GTCAATTTCTGATCCGAAGCAGTGGATACATGAAACACGCAAACTC 738
Qy 667 GATTATCTCGGTTGTGAAAGATACGACAAAAGATTGGTTTGTAAAGCAGCCGCTATCT 726
Db 739 GCAAAAGCAGAAATTTATGAATATTACTAGTATAA-----TATTTAAATCTCAATATAT 792
Qy 727 CAAGAATATGACAGATAAATGCGTGCAGGAAGTAACACCGGTGACAGGAAAGCGTTTAATG 786
Db 793 AAAGGAAACAAGATCTTTTGGCAACAAGATTTATGCCAGCTGTCAGAGGAGTTTAGAG 852
Qy 787 GAAGCGCAGCGGTTTACAGCTGTTTATATTTTGTGTTTGTGATACGTTATGT 839
Db 853 AAAGCGCAAGAAACACGCGCGGATTTATTTCAATTTATGTTTAAACATATGT 905

RESULT 8

US-10-421-654-97
; Sequence 97, Application US/10421654
; Publication No. US20040005604A1
; GENERAL INFORMATION:
; APPLICANT: Gramatikova, Svetlana
; APPLICANT: Hazlewood, Geoff
; APPLICANT: Lam, David E.
; APPLICANT: Barton, Nelson R.
; TITLE OF INVENTION: PHOSPHOLIPASES, NUCLEIC ACIDS ENCODING THEM AND
; TITLE OF INVENTION: METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-094001
; CURRENT APPLICATION NUMBER: US/10/421,654
; CURRENT FILING DATE: 2003-04-21
; PRIOR APPLICATION NUMBER: US 60/374,313
; PRIOR FILING DATE: 2002-04-19
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 97
; LENGTH: 1422
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-421-654-97

Query Match 10.1%; Score 85.8; DB 16; Length 1422;
Best Local Similarity 53.1%; Pred. No. 2e-10;
Matches 283; Conservative 0; Mismatches 232; Indels 18; Gaps 4;
Qy 313 TCTCACTTTTATGATCCGGATACCTGGAAACAACATATATTTCCCTTTTTCGAAACATGCAAAA 372
Db 385 TCTCATTTCTTACGATCCCGATACAGAAAGAACTATAAAGGGAGAGAGAACCAACAGCT 444
Qy 373 GAAAC---AGCGGCAAAATATTTTAACTTGTCTGCTCAAGCATACCAAAATCAAGATATG 429
Db 445 CTTTCTCAAGGAGATAAATATTTTAAATTTAGCAGGTGAATACTTTTAAAGAAAGATGATTGG 504
Qy 430 CAGCAAGCATTTCTTCTACTAGGATTTATCGCTTCAATTTTATAGGAGATGTTGAATCAGCCA 489
Db 505 AAACAAGCTTTCTATTTATTTAGGTGTCGACGCACTTACTTACAGATGCTACTCAGCCA 564
Qy 490 ATGCATGCGAGCAAACTTTTACGAATCT---TTCTTATCCAAATGGGTTTTCATTTCTAAATAC 546
Db 565 ATGCATGCTGCTAAATTTTACAGCTGTGACAGGAGTGTATATAAGCTTTTTCATAGCTTTT 624
Qy 547 GAAATTTTGTGATACAAATAAATAAATTAATATTTTGTTCAGATAGCAATGGATATTGG 606

625 GAAGATTATGACGACAAATTCAGAACAGTTTAAAGTATCAGATGAGAGGGAAATAT 684
 607 AATTGGAAAGAGCAAAACCCAGAGATTGGATTGAAGGAGCAGCGGTAGCAGCTAAACAA 666
 685 AATT-----TAGTAATCTTAATGATCCGAAACAGTGGATCCATGAACACAGCGAGACTC 738
 667 GATTATCTGCGGTTGGAAGATACGACAAAGATGGTTTGAAGAGCAGCGTATCT 726
 739 GCAAAAGTGAATTCGGGAACATTACCA-----ATGATGTGATTAAATCTCACTATAAT 792
 727 CAAGATATGCAATATGCGTGGCGGAAGTAAACACCGGTGACAGGAAGCGTTTAATG 786
 793 AAAGGAACAATGCTCTTTGGCAGCAAGAGTTATGCCAGCTGTTCAAGAGTTTAGAA 852
 787 GAAGCGAGCGCGTTTACAGCTGGTTTATATTCATTTGCTGGTTTGATACGTATGT 839
 853 CAAGCCCAAGAAATACGCGGGATTATTCATTTATGTTTAAACATATGT 905

RESULT 9
 US-10-421-654-89
 ; Sequence 89, Application US/10421654
 ; Publication No. US20040005604A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gramatikova, Svetlana
 ; APPLICANT: Hazlewood, Geoff
 ; APPLICANT: Lam, David E.
 ; APPLICANT: Barton, Nelson R.
 ; TITLE OF INVENTION: PHOSPHOLIPASES, NUCLEIC ACIDS ENCODING THEM AND
 ; FILE REFERENCE: 09010-094001
 ; CURRENT APPLICATION NUMBER: US 60/374,313
 ; PRIOR FILING DATE: 2003-04-21
 ; NUMBER OF SEQ ID NOS: 106
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 89
 ; LENGTH: 1422
 ; TYPE: DNA
 ; ORGANISM: Unknown
 ; FEATURE:
 ; OTHER INFORMATION: Obtained from an environmental sample
 US-10-421-654-89

Query Match 9.0%; Score 76.8; DB 16; Length 1422;
 Best Local Similarity 52.2%; Pred. No. 3.1e-08;
 Matches 280; Conservative 0; Mismatches 232; Indels 24; Gaps 4;

313 TCTCATTATGATCCGGATCTGGACAACTATATTCCTTTTTCGGAACATGCMAAA 372
 385 TCTCATTTTACGATCCCGATACGAGAAAGAACTATAAGGGGAGAGAACCAAGCT 444
 373 GAAAC---AGGCGCAAAATATTTAACTTGTCTGCTCAAGCATACCAAAATCAAGATG 429
 445 CTCTCTCAAGGAGATAAATTTTAAATTTAGCAGCGGATTTACTTTAAGAAAGAGGATTGG 504
 430 CAGCAGCATCTCTTACTTAGAGATTATCGCTTCAATTTATTTAGGAGATGTAATCAGCCA 489
 505 AAACAAGCTTTCTATTATTTAGGTTTGCAGCGATTTCTCAAGATGCTACTCAGCCA 564
 490 ATGCATGCTCTTCTTACTTAGAGATTATCGCTTCAATTTATTTAGGAGATGTAATCAGCCA 546
 565 ATGCATGCTCTTCTTACTTAGAGATTATCGCTTCAATTTATTTAGGAGATGTAATCAGCCA 624
 490 ATGCATGCTCTTCTTACTTAGAGATTATCGCTTCAATTTATTTAGGAGATGTAATCAGCCA 606
 565 ATGCATGCTCTTCTTACTTAGAGATTATCGCTTCAATTTATTTAGGAGATGTAATCAGCCA 684
 547 GAAATTTTGTGATACAAATAAATAATTAATTTATTTAGGAGATGTAATCAGCCA 663
 625 GAAATTTTGTGATACAAATAAATAATTAATTTATTTAGGAGATGTAATCAGCCA 744
 607 AATT---GGAAGAGCAAAACCCAGAGATTGGATTGAAGAGCAGCGGTAGCAGCTAAA 723
 685 AATTAGTGAATTTCTTAATGATCCAAACACAGTGGATCCATGAACACAGCGAGACTCGCAAAA 744

625 GAAGATTATGACGACAAATTCAGAACAGTTTAAAGTATCAGATGAGAGGGAAATAT 684
 607 AATTGGAAAGAGCAAAACCCAGAGATTGGATTGAAGGAGCAGCGGTAGCAGCTAAACAA 666
 685 AATT-----TAGTAATCTTAATGATCCGAAACAGTGGATCCATGAACACAGCGAGACTC 738
 667 GATTATCTGCGGTTGGAAGATACGACAAAGATGGTTTGAAGAGCAGCGTATCT 726
 739 GCAAAAGTGAATTCGGGAACATTACCA-----ATGATGTGATTAAATCTCACTATAAT 792
 727 CAAGATATGCAATATGCGTGGCGGAAGTAAACACCGGTGACAGGAAGCGTTTAATG 786
 793 AAAGGAACAATGCTCTTTGGCAGCAAGAGTTATGCCAGCTGTTCAAGAGTTTAGAA 852
 787 GAAGCGAGCGCGTTTACAGCTGGTTTATATTCATTTGCTGGTTTGATACGTATGT 839
 853 CAAGCCCAAGAAATACGCGGGATTATTCATTTATGTTTAAACATATGT 905

RESULT 10
 US-10-421-654-81
 ; Sequence 81, Application US/10421654
 ; Publication No. US20040005604A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gramatikova, Svetlana
 ; APPLICANT: Hazlewood, Geoff
 ; APPLICANT: Lam, David E.
 ; APPLICANT: Barton, Nelson R.
 ; TITLE OF INVENTION: PHOSPHOLIPASES, NUCLEIC ACIDS ENCODING THEM AND
 ; FILE REFERENCE: 09010-094001
 ; CURRENT APPLICATION NUMBER: US 60/374,313
 ; PRIOR FILING DATE: 2003-04-21
 ; NUMBER OF SEQ ID NOS: 106
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 81
 ; LENGTH: 1422
 ; TYPE: DNA
 ; ORGANISM: Unknown
 ; FEATURE:
 ; OTHER INFORMATION: Obtained from an environmental sample
 US-10-421-654-81

Query Match 8.5%; Score 72; DB 16; Length 1422;
 Best Local Similarity 51.7%; Pred. No. 4.5e-07;
 Matches 277; Conservative 0; Mismatches 235; Indels 24; Gaps 4;

313 TCTCATTATGATCCGGATCTGGACAACTATATTCCTTTTTCGGAAC---ATGCA 369
 385 TCTCATTTTACGATCCCGATACGAGAAAGAACTATAAGGGGAGAGAACCAACAGCT 444
 370 AAAGAAACAGCGCAAAATATTTAACTTGTCTGCTCAAGCATACCAAAATCAAGATG 429
 445 CTTCACAGGAGATAAATATTTTAAATTTAGCAGTGAATTTACTTTAGAGGGGCGCAAA 504
 430 CAGCAGCATTTCTTCTTACTTAGAGATTATCGCTTCAATTTATTTAGGAGATGTAATCAGCCA 489
 505 AAACAAGCTTTTATTATTTAGGTTTGCACAGCGATTTTACAGATGCTACTCAACCA 564
 490 ATGCATGCTCTTCTTACTTAGAGATTATCGCTTCAATTTATTTAGGAGATGTAATCAGCCA 546
 565 ATGCATGCTCTTCTTACTTAGAGATTATCGCTTCAATTTATTTAGGAGATGTAATCAGCCA 624
 547 GAAATTTTGTGATACAAATAAATAATTAATTTATTTAGGAGATGTAATCAGCCA 606
 625 GAAATTTTGTGATACAAATAAATAATTAATTTATTTAGGAGATGTAATCAGCCA 684
 607 AATT---GGAAGAGCAAAACCCAGAGATTGGATTGAAGAGCAGCGGTAGCAGCTAAA 663
 685 AATTAGTGAATTTCTTAATGATCCAAACACAGTGGATCCATGAACACAGCGAGACTCGCAAAA 744
 664 CAAGATATCTCGGCGTTGTGAAACAGTACGACAAAGATTTGGTTTGTAAAGCAGCGCTA 723
 745 GTGGAATCGGGAACATTACCAATGACGAGATTAATCTTCACTAATAAAGGAAACAAAT 804
 724 TCTCAAGATATGACAGATAAATGCGGTGCGGAAGTAAACACCGGTGACAGGAAGCGTTTA 783

Db 805 GTC-----TTTGGCAACAAGATTATGCCAGCTGTCAGAGGAGTTTA 849
QY 784 ATGGAAGCGCAGCGGTTACAGCTGTTATATTTCTATTTGTTGTTGATACGATGT 839
Db 850 GAGACGCAACAAGAAACACGCGGGATTATTCATTATGTTTAAACATTGT 905

RESULT 11

US-10-398-221-3772/c
; Sequence 3772, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3772
; LENGTH: 4562
; TYPE: DNA
; ORGANISM: Listeria monocytogenes 4b
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u

US-10-398-221-3772

Query Match 6.8%; Score 57.4; DB 16; Length 4562;

Best Local Similarity 53.3%; Pred. No. 0.0029; Mismatches 106; Indels 0; Gaps 0;
Matches 121; Conservative 0;

QY 107 TAAAGATGGTCAGCTGAGGATAAGCAATAATGAGGGATTAACTCTCATTTGTGGATTGTA 166
Db 227 TTAGTTGGTCCGGGTAATCCGACAAATCTGACGTAATAGCACATATGCTTTT 168
QY 167 ATCGTGCAATTGACATCATGTCTCGTAATACACGATTGTGAATCGGAATGAACATGCA 226
Db 167 AACAGCAGAAAAAATCTAGCTAAAGATGTAGATCATATGCGAGCTAATTTAATGAATG 108
QY 227 TATTAAATGAGTGGCGTGTGATTAGAAAAATGTTATTTCTGCTGATTACGAGATC 286
Db 107 AACTTAAAAATTCGACAAACAAATTTGCTCAAGGAATATATGACGCGGATCATAAAATC 48
QY 287 CTTATTATGATAATAGTACATATGCTTCTCAGTTTATGATCCGGAT 333
Db 47 CATATTATGATAGTACGTTTATCTCATTTTATTAATCTGAT 1

RESULT 12

US-08-781-986A-470
; Sequence 470, Application US/08781986A
; Publication No. US2003005436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 470:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9821 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-781-986A-470

Query Match 5.1%; Score 43.4; DB 8; Length 9821;

Best Local Similarity 44.3%; Pred. No. 11;
Matches 176; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

QY 321 TTATGATCGGATACCTGGAAACAACATATATCTCTTTCGAAACATGCAAAAGAACAGC 380
Db 4223 TGATGTCAGATTTCAGAAATGAAGAAAGAAATTTGAAAGCTTGAANATCAAGAGC 4282
QY 381 CGCAAAATATTTTAACTTGTCTGTCGTCACAGCATACCAAAATCAAGATATGCAGCAAGCAT 440
Db 4283 TGACATCACTGTTAAAGATTGAGTAAGCTGAAAAGAGCAAGAGCGTATTTTAGTAAG 4342
QY 441 CTTCTACTTAGGATTTATCGCTTCATTTATTTAGGAGATGTGAATCAGCCATCGATCGACG 500
Db 4343 AATGCAAGAAACAGAAATGCTTTATTTCAATAGCAGAGCGAGCAAGCAATTTAAAGAGC 4402
QY 501 AAACCTTTACGATCTTTCTATCCATGGTTTCCATTTCAATACGAAATTTTGTTCGA 560
Db 4403 AGAAAAAGCAAGAAAGCAAGAAAAAGAAAGTGAATAGCAGTATGAAGATGATGTCT 4462
QY 561 TACAATAAAAAATACTATATTTGTTTCAGATAGCAATGGATTTGGAATTTGAAAGGAGC 620
Db 4463 TGCTATAAAAAATAAGCTCAACCTTTCTAAGCTCGAAAAAGATAAATTTGTTAGCTATTGC 4522
QY 621 AAACCCAGAAGTTGGATTGAAGAGCGAGCGGTAGCAGCTAAACAGATTTATCTGGCGT 680
Db 4523 TGATCAAGACATAGAGTGAAGTAAGAAAGGCAAAATCTAAAAAAGATGCTGTAGTAGA 4582
QY 681 TGTGAACGATACGACAAAGATTGGTTTGTAAAGCA 717
Db 4583 CGTTGTTAAAAAGCAAAATAAGATATTGATAAGAA 4619

RESULT 13

US-10-329-624-470
; Sequence 470, Application US/10329624
; Publication No. US20040040307A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue


```
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/329,624
; FILING DATE: 27-Dec-2002
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/956,171
; FILING DATE: October 20, 1997
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248PLD1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
;
; INFORMATION FOR SEQ ID NO: 470:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9821 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 470:
US-10-329-624-470

Query Match 5.1%; Score 43.4; DB 13; Length 9821;
Best Local Similarity 44.3%; Pred. No. 11;
Matches 176; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

QY 321 TTATGATCGGATCTGGACACATATATCTCTTTGGAAACATGCAGAAACAGG 380
DB 4223 TGATGTCAGATTCAGAAAATGNAAGAAAGAAATTTGAAAAGCTTGAAAATCAAAGACG 4282

QY 381 CGCAAAATATTTAACTTGGTCAAGCATACCAAAATCAAGATATGCAGCAAGCAT 440
DB 4283 TGACATCAGTTTAAAGATTGAGTAAGACTGAAAAGAGCAGACGCTATTTTAGTAAG 4342

QY 441 CTCTACTTAGGATTATCGCTTCATTTTAGGAGATGGAATCAGCCAAATGCATGCAGC 500
DB 4343 AATGCAAGAAACAGAAATGCTTATTCATAGCAGGAGCGAGCAAGCAATTAAGAAGC 4402

QY 501 AAATTTAGCAATCTTTCTATCCATGGGTTTCCATTTAAATACGAAAATTTTGTGA 560
DB 4403 AGAAAAGCAAGAAAGCAAGAAAGAAAGAAAGTGGATAAGCAGTATGAAGATGATGTCAT 4462

QY 561 TACAATAAAAATAACTATATTTTTCAGATAGCAATGATATTTGAAATTTGAAAAGGAGC 620
DB 4463 TCGTATAAAAAATAGCTCAACCTTTCTAAGTCTGAAAAGATAAATTTAGCTATTGC 4522

QY 621 AAACCCAGAAATTTGGAATTTGAAGGAGCAGCGGTAGCAGTAAACAGATTTATCTGGCGCT 680
DB 4523 TGATCAAGACATAGGATGAAAGTAAAGAAAGGCAAAATCTAAAAAAGATGCTGTAGTAGA 4582

QY 681 TGTAACGATACGACAAAGATTGGTTTCTTAAAGCA 717
DB 4583 CGTTGTTAAAGCAAAATAAGATATTGATAAGAA 4619

RESULT 14
US-09-966-521-21/c
; Sequence 21, Application US/09966521
; Publication No. US20030087321A1
; GENERAL INFORMATION:
; APPLICANT: TOMICH, Che-Shen
; APPLICANT: QUINN, Cheryl
; APPLICANT: ARVIDSON, Staffan
; APPLICANT: HARRIS, Douglas
; APPLICANT: MOTT, John
; TITLE OF INVENTION: ANTIMICROBIAL METHODS AND MATERIALS
; FILE REFERENCE: 00774.US1 CN1
; CURRENT APPLICATION NUMBER: US/10/429,094
; CURRENT FILING DATE: 2003-05-02
; PRIOR APPLICATION NUMBER: US 09/966,521
; PRIOR FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 837
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-10-429-094-21

Query Match 5.1%; Score 43.2; DB 15; Length 837;
Best Local Similarity 50.5%; Pred. No. 3.4;
Matches 105; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 403 GGTCAAGCATACCAAAATCAAGATATGCAGCAAGCATTTCTTACTTAGGATTATCGCTT 462
DB 820 GGCCAATCGACCTTGACCAAGTATGTCAGAAATAACTGGTGTAGTTACATTATATCGT 761

QY 463 CATTATTTAGGAGATGTAATCAGCCAAATGCATGCAGCAAACTTTTACGAAATCTTTCTTAT 522
DB 760 AATTATTCACATGAAAAGCTTCATTAATAAACTTTCTTCAATTTATCAACATATTCGAATGA 701

QY 523 CCAATGGGTTTCCATTTCTAATACGAAAATTTTGTGTATACATAAATAAATACTATAT 582
DB 700 CGTTAGCATGTGCGACACCAATGGATTTGATTTTCATGATCTCTATAAAATTCAGCAATTT 641

QY 583 GTTTCAGATAGCAATGGATATTTGGAATT 610
DB 640 CTTTTCAGTATTTGATAGTACTAGAAATT 613

RESULT 15
US-10-429-094-21/c
; Sequence 21, Application US/10429094
; Publication No. US20030180821A1
; GENERAL INFORMATION:
; APPLICANT: TOMICH, Che-Shen
; APPLICANT: QUINN, Cheryl
; APPLICANT: ARVIDSON, Staffan
; APPLICANT: HARRIS, Douglas
; APPLICANT: MOTT, John
; TITLE OF INVENTION: ANTIMICROBIAL METHODS AND MATERIALS
; FILE REFERENCE: 00774.US1 CN1
; CURRENT APPLICATION NUMBER: US/10/429,094
; CURRENT FILING DATE: 2003-05-02
; PRIOR APPLICATION NUMBER: US 09/966,521
; PRIOR FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 837
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-10-429-094-21

Query Match 5.1%; Score 43.2; DB 15; Length 837;
Best Local Similarity 50.5%; Pred. No. 3.4;
Matches 105; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 403 GGTCAAGCATACCAAAATCAAGATATGCAGCAAGCATTTCTTACTTAGGATTATCGCTT 462
DB 820 GGCCAATCGACCTTGACCAAGTATGTCAGAAATAACTGGTGTAGTTACATTATATCGT 761

QY 463 CATTATTTAGGAGATGTAATCAGCCAAATGCATGCAGCAAACTTTTACGAAATCTTTCTTAT 522
DB 760 AATTATTCACATGAAAAGCTTCATTAATAAACTTTCTTCAATTTATCAACATATTCGAATGA 701
```

Qy	523	CCAATGGGTTCCATTCTAAATACGAAATTTTGTGTGATACAAATAAAAAATAACTATATT	582
Db	700	CGTTAGCATGTGCGACACCAATGGATTGATTTTCATGATCTCCTATAAAATTCAGCAATT	641
Qy	583	GTTCAGATAGCAATGGATATTGGAATT	610
Db	640	CCTTTTCAAGTATTGGATACTAGAATT	613

Search completed: August 2, 2004, 04:20:53
Job time : 511 secs

This page blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 1, 2004, 23:39:34 ; Search time 432 Seconds
(without alignments)
8348.889 Million cell updates/sec

Title: PCT-US03-12556-1
Perfect score: 849
Sequence: 1 atgaaaagaagattagc.....atcgtatgtaaatcgctaa 849

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:.*
1: Geneseq1980s:.*
2: Geneseq1990s:.*
3: Geneseq2000s:.*
4: Geneseq2001as:.*
5: Geneseq2001bs:.*
6: Geneseq2002s:.*
7: Geneseq2003as:.*
8: Geneseq2003bs:.*
9: Geneseq2003cs:.*
10: Geneseq2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	154.4	18.2	870	6	ABQ67927 Listeria
2	154.4	18.2	870	6	ABQ69833 Listeria
3	154.4	18.2	110000	6	Continuation (3 of
C 4	57.4	6.8	4562	6	ABQ70959 Listeria
C 5	46.6	5.5	3706	2	AAT36392 Lactococc
6	45.8	5.4	9821	2	AAV74781 Staphyloc
7	45.4	5.3	6033	3	AAA70152 Plasmodiu
8	45.2	5.3	3738	3	AAA70178 Plasmodiu
9	45	5.3	2000	7	ADA71938 Rice gene
10	43.4	5.1	6222	7	ACF72350 Staphyloc
C 11	43.2	5.1	837	7	ACF74582 Staphyloc
C 12	43.2	5.1	837	9	ADC21339 Staphyloc
C 13	43.2	5.1	837	9	ADC25025 DNA encod
C 14	43.2	5.1	837	9	ADD52527 Staphyloc
15	43.2	5.1	2700	3	AAA70220 Plasmodiu
C 16	43.2	5.1	5924	2	AAV74441 Staphyloc
C 17	43	5.1	8170	6	ABK28258 DNA trans
18	43	5.1	43095	3	AAA68254 Bacteriop
C 19	42	4.9	969	6	ABQ39490 Oligonucle
C 20	42	4.9	969	6	ABQ39491 Oligonucle
C 21	41.8	4.9	6227	6	ABL33586 Human imm
C 22	41.8	4.9	6227	6	ABL92260 Chemical
23	41.6	4.9	2319	4	AAC84742 S. pneumo

24	41.6	4.9	2911	2	AAV52332 Streptoco
C 25	41.6	4.9	3840	7	ABQ77399 Human col
26	41.6	4.9	6361	3	AAZ91806 Streptoco
27	41.6	4.9	6420	7	ABX06302 S. pneumo
28	41.6	4.9	6423	7	ACA49616 Prokaryot
C 29	41.6	4.9	110000	6	Continuation (17 o
30	41.6	4.9	110000	6	ABQ67195 Listeria
31	41.6	4.9	110000	7	Continuation (5 of
32	41.6	4.9	110000	7	Continuation (6 of
C 33	41.2	4.9	2000	7	ADA71938 Rice gene
34	41	4.8	3491	4	ABL19460 Drosophil
C 35	41	4.8	4985	6	ABQ75107 Anopheles
C 36	41	4.8	4985	9	ACF79720 Mosquito
37	40.8	4.8	3837	3	AAV70211 Plasmodiu
C 38	40.2	4.7	1014	7	ACA43139 Prokaryot
39	40	4.7	3381	4	AAZ51736 Staphyloc
40	40	4.7	3477	4	AAZ54787 Staphyloc
41	39.8	4.7	1113	2	AAQ62305 Clostridi
C 42	39.8	4.7	1353	2	AAZ99537 Nucleic a
43	39.8	4.7	2130	7	ACA52696 Prokaryot
44	39.4	4.6	705	2	AAQ10244 Streptoco
45	39.4	4.6	1617	2	AAT49317 Type-6 M-

ALIGNMENTS

RESULT 1
ABQ67927
ID ABQ67927 standard; DNA; 870 BP.
XX
AC ABQ67927;
XX
DT 29-AUG-2003 (revised)
DT 29-AUG-2002 (first entry)
DE Listeria monocytogenes EGD DNA sequence #51.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX
OS Listeria monocytogenes; EGD.
XX
PN WO200228891-A2;
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR003061.
XX
PR 04-OCT-2000; 2000FR-00012697.
XX
PA (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators.
XX
PS Claim 16; SEQ ID NO 740; 180pp; French.
XX
CC The present invention relates to nucleic acid sequences (ABQ67188-
CC ABQ71212) from Listeria sp. The sequences are useful as probes and
CC primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of gene
CC expression. Proteins encoded by the nucleic acid sequences can be used to
CC screen for compounds that modulate gene expression, replication and
CC pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in anti-
CC Listeria vaccines. Note: The sequence data for this patent did not form


```
Qy 638 TTGAAGGAGCGGTAGCAGCTAAACAAGATTATCTCGGCGTTGTGAACGATACGACAA 697
Db 680 TCTATGAAAATGCGAAAAGGCGAAGCGGACTACCCGAAAATAGTCAATGCGAAAACCTA 739

Qy 698 AAGATTGGCTTCTTAAAGCA 717
Db 740 AAAAATCATATTTAGTAGGA 759

RESULT 3
ABAO3041_02
Continuation (3 of 30) of ABA03041 from base 200001 (Listeria monocytogenes EGD-e genome
WP Sequence split into 30 fragments LOCUS ABA03041 Accession ABA03041
WP Fragment Name Begin End
WP ABA03041_00 1 110000
WP ABA03041_01 100001 210000
WP ABA03041_02 200001 310000
WP ABA03041_03 300001 410000
WP ABA03041_04 400001 510000
WP ABA03041_05 500001 610000
WP ABA03041_06 600001 710000
WP ABA03041_07 700001 810000
WP ABA03041_08 800001 910000
WP ABA03041_09 900001 1010000
WP ABA03041_10 1000001 1110000
WP ABA03041_11 1100001 1210000
WP ABA03041_12 1200001 1310000
WP ABA03041_13 1300001 1410000
WP ABA03041_14 1400001 1510000
WP ABA03041_15 1500001 1610000
WP ABA03041_16 1600001 1710000
WP ABA03041_17 1700001 1810000
WP ABA03041_18 1800001 1910000
WP ABA03041_19 1900001 2010000
WP ABA03041_20 2000001 2110000
WP ABA03041_21 2100001 2210000
WP ABA03041_22 2200001 2310000
WP ABA03041_23 2300001 2410000
WP ABA03041_24 2400001 2510000
WP ABA03041_25 2500001 2610000
WP ABA03041_26 2600001 2710000
WP ABA03041_27 2700001 2810000
WP ABA03041_28 2800001 2910000
WP ABA03041_29 2900001 2944528

Query Match 18.2%; Score 154.4; DB 6; Length 110000;
Best Local Similarity 53.1%; Pred. No. 1.6e-28;
Matches 329; Conservative 0; Mismatches 291; Indels 0; Gaps 0;

Qy 98 CACCGATTTTAAGATGGTCAGCTGAGGATAGCATATGAGGGGATTAACTCTCATTTGT 157
Db 11564 CACATAAACTTAGTTGGTCCGGGATACCCGACAAATCTGACGTAATAGGCACTATT 11623

Qy 158 GGATTGTAATCGTGCATATGATCTCTCGTAATCAACGATGTGTAATCGGAATG 217
Db 11624 GGCTTTTAAACAGCGGAAAAAATAGTCTAAAGATGTAATCATATGCGAGCTAAT 11683

Qy 218 AAATGTCATTATTAATGATGGCGTGTGATTAGAAATGGTATTATTTCTGCTGATT 277
Db 11684 TAATGAATGAACTTAAAAAATTCGATAAACAATAGCTCAAGGAATATATGATCGGATC 11743

Qy 278 ACGAGATCCTTATTATGATATAGTACATATGCTTCTCACCTTTTATGATCGGATACG 337
Db 11744 ATAAAAATCCATATTATGATAGTACTAGTACATTTTATCTCAATTTTATATCTCTGATAG 11803

Qy 338 GAACAACATATATCTCTTTTGGAAACATGCAAAAGAACAGCGCCAAATATTTAAACC 397
Db 11804 ATATATCTTATTGGCGGTTTGTCTAATGCGAAAATAACAGGACAAAGATTATTTCAATC 11863

Qy 398 TTGCTGGTCAAGCATACCAAAATCAAGATATGACGACGATCTTCTTACTTAGGATTAT 457
Db 11864 AATCGGTGACTGATTACCGGAGAGGAAATTTGACACGCGTTTATATAAATTAGGCCTAG 11923
```

```
Qy 458 CGCTTCATTATTATTAGGAGATGTGAATCAGCCAATGCAATGCAAGAACTTTTACGAATCTTT 517
Db 11924 CAATCCATTATTATACGGATATTAGTCAACTATGCAAGCCAATTAATTTTACCGCAATAT 11983

Qy 518 CTTATCCAAATGGTTTCCATTCTAAATACGAAAATTTTGTGATACATAAAAATAACT 577
Db 11984 CATACCTCCAGGCTACCACTGTGCATATGAAAATTTACGTAGTACCACTTAAACACAAT 12043

Qy 578 ATATTGTTTTCAGATAGCAATGGAATTTGGAATTTGAAAAGGAGCAAAACCCAGAAAGATTGGA 637
Db 12044 ATCAAGCAACGAGACATGTTAGCAAAAAGATTTTGTCTCAGATGACGTGAAGACTGGC 12103

Qy 638 TTGAAGGAGCGGTAGCAGCTAAACAAGATTATCTCGGCGTTGTGAACGATACGACAA 697
Db 12104 TCTATGAAAATGCGAAAAGGCGAAGCGGACTACCCGAAAATAGTCAATGCGAAAACCTA 12163

Qy 698 AAGATTGGTTGTTAAAGCA 717
Db 12164 AAAAATCATATTTAGTAGGA 12183

RESULT 4
ABQ70959/C
ID ABQ70959 standard; DNA; 4562 BP.
XX
AC ABQ70959;
XX
DT 29-AUG-2003 (revised)
DT 29-AUG-2002 (first entry)
XX
DE Listeria monocytogenes 4b contig DNA sequence #901.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX
OS Listeria monocytogenes ATCC 19115.
XX
WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR003061.
XX
PR 04-OCT-2000; 2000FR-00012697.
XX
PA (INSP ) INST PASTEUR.
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX
WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators.
XX
PS Claim 14; SEQ ID NO 3772; 180pp; French.
XX
CC The present invention relates to nucleic acid sequences (ABQ67188-
CC ABQ71212) from Listeria sp. The sequences are useful as probes and
CC primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of gene
CC expression. Proteins encoded by the nucleic acid sequences can be used to
CC screen for compounds that modulate gene expression, replication and
CC pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in anti-
CC Listeria vaccines. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. (Updated
CC on 29-AUG-2003 to standardise OS field)
XX
SQ Sequence 4562 BP; 1166 A; 943 C; 782 G; 1671 T; 0 U; 0 Other;
```

Query Match 6.8%; Score 57.4; DB 6; Length 4562;
 Best Local Similarity 53.3%; Pred. No. 0.00024;
 Matches 121; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

Qy 107 TAAGATGCTGAGTGAAGCAATATGAGGGGATTAACCTCTATTGTGGATTCTAA 166
 Db 227 TTATTTGCTCGCGAATAATCCGCAATATGACGTAAATACGCACTATTGGCTTTTA 168

Qy 167 ATCGTGAATTTGACATCATGCTCTGTAATACACGATTTGGAATCCGAATGAATCTGCAT 226
 Db 167 AACAGCAGAAAAAATACTAGCTAAAGATGATGATCATATCGAGCTAAATTAATGAATG 108

Qy 227 TATTAAATGAGTGGCTGCTGATTAGAAATGATTTATTTCTCTGATTACGAGATC 286
 Db 107 AACTTAAATTTTCGACAAACAAATGCTCAAGAAATATATGACGCGGATCAATAAATC 48

Qy 287 CTTATTATGATAATGATACATATGCTCTCACTTTTATGATCCGAT 333
 Db 47 CATATTATGATACTAGTACGTTTTTAATCTCATTTTATATCTCTGAT 1

RESULT 5
 AAT36392/c
 ID AAT36392 standard; DNA; 3706 BP.

XX AAT36392;
 AC
 XX

DT 16-OCT-2003 (revised)
 DT 08-DEC-1996 (first entry)

XX Lactococcus lactis subsp. cremoris W9 restriction-modification system.
 XX
 XX Lactic acid bacterium; Danish starter culture; cheese, enzyme;
 KW restriction endonuclease; methylase; fermented milk; phage resistance;
 KW ds.
 XX

OS Lactococcus lactis subsp. cremoris; W9.
 XX
 XX Key Location/Qualifiers
 FH complement(422..2161)
 CDS /tag= a
 FT /codon_start= 422
 FT /product= "LlaBI methylase"
 FT /evidence= EXPERIMENTAL
 FT /gene= "ORF"
 FT /number= 1 /standard name= "Gene coding for LlaBI
 FT methylase
 FT /label= m-llaBI
 FT 2464..3360
 FT /tag= b
 FT /codon_start= 2464
 FT /product= "LlaBI restriction endonuclease"
 FT /evidence= EXPERIMENTAL
 FT /gene= "ORF"
 FT /number= 2 /standard name= "Gene coding for LlaBI
 FT endonuclease
 FT /label= r-llaBI

XX WO9625503-A1.
 XX
 XX 22-AUG-1996.
 XX
 XX 19-FEB-1996; 96WO-DK000076.
 XX
 XX 17-FEB-1995; 95DK-00000179.
 XX
 XX (JOSE/) JOSEPHSEN J.
 XX (NYEN/) NYENGAARD N. R.
 XX (VOGE/) VOGENSEN F. K.
 XX (MAUS/) MADSEN A.
 XX Josephsen J, Nyengaard NR, Vogensen FK, Madsen A;

XX WPI; 1996-393404/39.
 DR P-PSDB; AAW02164, AAW02165.
 XX
 XX Type II restriction-modification systems obtd. from Lactococcus lactis -
 PT for conferring phage resistance on lactic acid bacteria, useful as
 PT starter cultures for cheese and fermented milk prods.
 XX
 XX Claim 2; Page 67-69; 93pp; English.
 XX
 XX This DNA encodes the plasmid-derived type-II restriction- modification
 CC system from Lactococcus lactis subsp. cremoris W9, llaBI, and contains 2
 CC ORFs transcribed in the same direction and coding for putative proteins
 CC of 580 AA (llaBI methylase M.LlaBI; AAW02164) and 299 AA (llaBI
 CC endonuclease R.LlaBI; AAW02165). This restriction-modification system may
 CC be used in a method for conferring increased virus resistance, more
 CC specifically phage resistance, to a L. lactis strain used in cheese
 CC manufacture. (Updated on 16-OCT-2003 to standardise OS field)
 XX
 XX Sequence 3706 BP; 1284 A; 535 C; 575 G; 1312 T; 0 U; 0 Other;
 SQ

Query Match 5.5%; Score 46.6; DB 2; Length 3706;
 Best Local Similarity 47.7%; Pred. No. 0.12; 149; Indels 0; Gaps 0;
 Matches 136; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

Qy 51 AGTTCAAAGTGTAGTATTTCACAAACAAATAATAGTGAAGTCTCCACCGATTTTAAAG 110
 Db 1010 ATTATATCGAATAATATATCTTATGAATAATATGAACGCAATGCTGTACCAATCAGCAT 951

Qy 111 ATGCTCAGCTGAGGATTAAGCATTAATGAGGGGATTAACCTCATTTGTGGATTGTAATCG 170
 Db 950 TGTCTCAATGAAGGAAAAATTCCTTAATACCTTTAAATTAATTTATTTATGCTATTAAAGACGA 891

Qy 171 TGCAATTGACATCATGCTCTGTAATACAAAGATTGTGAATCCGAATGAACTGCATTATT 230
 Db 890 GCTCAGTAAACGCAATGCTGTGGAACCAACCGTTAACATTTGGTATGAATATGGCGATC 831

Qy 231 AAATGAGTGGCGTCTGATTTAGAAAAATGGTATTATTCTGCTGATACGGAATCCCTTA 290
 Db 830 TCAAGCATTTGAATAGCCATGTACCTAAATAATTTTCTTACAAATTTCTTTGAATCCAAA 771

Qy 291 TTATGATAATAGTACATATGCTTCTCATTCTTATGATCCGGATAC 335
 Db 770 CTTTGTATATTTTACAGATTATGCTCTTTTAAATTAATGGATATGC 726

RESULT 6
 AAV74781
 ID AAV74781 standard; DNA; 9821 BP.
 XX
 XX AAV74781;
 AC
 XX
 XX 16-MAR-1999 (first entry)
 DT
 XX
 XX DE
 XX Staphylococcus aureus contig SEQ ID #470.
 XX
 XX Computer readable medium; vaccine; S.aureus infection; immunodetection;
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.
 XX
 XX Staphylococcus aureus.
 OS
 XX
 XX Key Location/Qualifiers
 FH misc_feature 1021..1080
 FT /tag= a
 FT /note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. they are
 FT included to maintain the nucleotide numbering given in
 FT the specification for this DNA sequence"
 FT 2821..2880
 FT misc_feature /tag= b
 FT /note= "these bases represent a line of missing text in
 FT

the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

4621..4680

/*tag= c

/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

6421..6480

/*tag= d

/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

8221..8280

/*tag= e

/note= "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

EP786519-A2.

30-JUL-1997.

07-JAN-1997; 97EP-00100117.

05-JAN-1996; 96US-0009861P.

(HUMA-) HUMAN GENOME SCI INC.

Kunsch CA, Choi GH, Barash SC, Dillon PJ, Fannon MR, Rosen CA; WPI; 1997-374922/35.

Polynucleotide(s) and proteins derived from *Staphylococcus aureus* - stored on computer readable medium and used in the production of anti-S.aureus vaccines.

Claim 1; Page 1387-1392; 3271pp; English.

This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences of the invention. The DNA sequences are recorded on a computer readable medium, preferably selected from a floppy or hard disk, random access memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using the S.aureus DNA sequences allows putative functions to be assigned so that protein-encoding or regulatory regions of commercial, therapeutic or industrial importance can be obtained. Specifically, sequences which are likely to encode antigens have been identified and these polypeptides can be used in a vaccine composition against S.aureus infection. The polypeptides can also be used in a kit for the immunodetection of S.aureus in a sample. S.aureus is implicated in numerous human diseases, including cellulitis, eyelid infections, food poisoning, osteomyelitis, skin and surgical wound infections, scalded skin syndrome, toxic shock syndrome, etc. Organisms transformed with the DNA sequences can be used for recombinant production of the polypeptides. The new DNA sequences (and their fragments) are useful as primers or probes for isolating homologues of any of the S.aureus DNA sequences contained on the computer readable medium

Sequence 9821 BP; 3652 A; 1311 C; 1978 G; 2575 T; 0 U; 305 Other;

Query Match 5.4%; Score 45.8; DB 2; Length 9821;

Best Local Similarity 38.8%; Pred. No. 0.26;

Matches 179; Conservative 0; Mismatches 282; Indels 0; Gaps 0;

Qy 321 TTATGATCCGGATCTGGAACAACATATATTCCTTTTCGGAACATGCAAAAGAAACAGG 380

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 4223 TGATGGTCAGATTTCAGAAATGAAGAAAGAAATTCGAAAGCTTGAATCAAGACG 4282

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 381 CGCAAAATATTTTAACTTCTGGTCAAGCATACCAAAATCAAGATATGCAGCAGCAATT 440

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 4283 TGACATCACTGTTAAAGAAATTGAGTAAGACTGAAAAAGAGCAAGCGTATTTTAGTAG 4342

Qy 441 CTTCTACTTGAAGATTATCGCTTTCAATTTATTAGGAGATGTGAATCAGCCAATGCATGCAGC 500

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 4343 AATGCAAGAAACAGAAATGCTTATTCAATAGACGAAGCGCAAGCAATTTAAAGAACG 4402

Qy 501 AAACTTTACGAATCTTTCTTATCCAAATGGTTTCAATTCCTAAATACGAAAAATTTTGTGGA 560

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 4403 AGAAAAACCAAGAAAGCAAGAAAAAGAAAGTGGTAAGCAGTATGAAGATGATGTCTAT 4462

Qy 561 TACAAATAAAAATAACTATATTGTTTTCAGATAGCAATGATGATATTCGAATTTGAAAGGAGC 620

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 4463 TGTATATAAAATTAACGTCAACCTTTCTAAGTCTGAAAAAGATAAATTTGTAGCTATTGTC 4522

Qy 621 AAACCCAGAGATTGGATTGAAGGAGCAGCGGTAGCAGCTAAACAAAGATTATCTCTGGCGT 680

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 4523 TGATCAAGACATTAAGGATGAAGTAAGAAAGCAAAATCTTAAAGAGATGCTGTAGTAGA 4582

Qy 681 TGTGAACGATAGCAGCAAAAGATTGGTTTGTAAAGCAGCGGTATCTCAAGATATATGCAGA 740

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 4583 CGTTGTTTAAAAAGCAAAATAAAGATATTTGATAAAGAAAAANNNNNNNNNNNNNNNNN 4642

Qy 741 TAAATGGCGTGGGAAGTAACACCGGTGACAGGAAAGCGTT 781

|||||

Db 4643 NNN 4683

RESULT 7

AAA70152

ID AAA70152 standard; DNA; 6033 BP.

XX

AC AAA70152;

XX

DT 07-NOV-2000 (first entry)

XX

DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:285.

XX

KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine; antimalarial; malaria; protozoacide; infection; insecticide; ds.

KX

OS Plasmodium falciparum.

XX

FN WO200025728-A2.

XX

PD 11-MAY-2000.

XX

PF 05-NOV-1999; 99WO-US026796.

XX

PR 05-NOV-1998; 98US-0107131P.

XX

PA (HOFF/) HOFFMAN S.

PA (CARU/) CARUCCI D.

PA (GARD/) GARDNER M.

XX (VENT/) VENTER J C.

PI Hoffman S, Carucci D, Gardner M, Venter JC;

DR WPI; 2000-365347/31.

XX

PT Proteins encoded by chromosome 2 of the human malarial parasite, Plasmodium falciparum, useful as antimalarial vaccines and in the diagnosis of P.falciparum infection.

PT

XX Disclosure; Page 493-495; 577pp; English.

XX

CC The present invention describes proteins and their fragments (I) encoded by chromosome 2 of the human malarial parasite, Plasmodium falciparum. CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2) CC vaccines against P. falciparum infection comprising (I) or (II). (I) and CC (II) are useful for the development of vaccines against P. falciparum CC infection. (1) and polyclonal antisera or a monoclonal antibody raised to CC immunogens comprising the sequences of (I), are useful in the detection CC of infection with P. falciparum. Furthermore, (I) (especially when they

are xifins or secreted or membrane proteins) can aid the identification of drugs to treat or prevent *P. falciparum* infection, or they can be used to identify drug resistance in *P. falciparum*. Sequencing of the Plasmodium chromosome 2 and the subsequent identification of proteins encoded by it will help to expand our understanding of parasite biology, a process hampered by the complexity of the parasitic lifecycle, and provide new targets for vaccine and drug development. Parasite resistance to drugs and mosquito resistance to insecticides have led to a resurgence of malaria in many parts of the world, and there is a pressing need for vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide and protein sequences given in the present invention, but which are not specifically mentioned within the specification

XX Sequence 6033 BP; 3019 A; 437 C; 707 G; 1870 T; 0 U; 0 Other;

Query Match 5.3%; Score 45.4; DB 3; Length 6033;
Best Local Similarity 45.2%; Pred. No. 0.29;
Matches 166; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 257 ATGGTATTATTCTGCTGATTACGAGATCCTTATTATGATTAATAGTACATATGCTTCTC 316

DB 5042 ATAAATATTTTATGATATCATCATGATACCAATAATAATTAATTAATAATAATAATA 5101

QY 317 ACTTTATGATCGGATACGGAACACATATATTCCTTTTGGAAACATGCAAAAGAAA 376

DB 5102 ATAATAATAGTAATAATAATAATAATAATATATGCTTAAGAAATAATAATAATA 5161

QY 377 CAGGCGCAAAATATTTTACCTTGCTGTCAGCATACCAAAATCAAGATATGCGAAG 436

DB 5162 TAATGATGAGATATCAATGCAAAATAAAGAGAAAGCTTAAAAAATAAAAAAAAA 5221

QY 437 CATTCCTCTACTAGGATATCGCTTCATTATTAGGAGATGTAATCAGCCAAATGCATG 496

DB 5222 AAAAAAATTTGTATACAAAAAATAATAATATTGTGGAAGGAAAAAATCAATATTC 5281

QY 497 CAGCAAACTTACGAATCTTTCTTATCCAAATGGGTTTCATTTCAATAACGAAAAATTTG 556

DB 5282 AATAATAATAGTTTCAAGTACATATTTTAATACCTGTGAGATTTCTTTAAAAATGAAAGATATTG 5341

QY 557 TTGATACAAATAAAAAATACTATATTTGTTTCAGATAGCAATGGATATTGGAAATGGAAAG 616

DB 5342 CTAAATAATAACAAATAAAGTGTGATGAATAAGATGATATCATGTTAATAATACATGA 5401

QY 617 GAGCAAA 623

DB 5402 GAGAAAA 5408

RESULT 8

ADA70178

ID AAA70178 standard; DNA; 3738 BP.

XX AAA70178;

AC AAA70178;

XX 07-NOV-2000 (first entry)

DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:311.

XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX antimalarial; malaria; protozoacide; infection; insecticide; ds.

OS Plasmodium falciparum.

PN WO200025728-A2.

XX 11-MAY-2000.

XX 05-NOV-1999; 99WO-US026796.

XX 05-NOV-1998; 98US-0107131P.

XX (HOFF/) HOFFMAN S.

PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
XX (VENT/) VENTER J C.

PI Hoffman S, Carucci D, Gardner M, Venter JC;

XX WPI; 2000-365347/31.

DR Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of *P. falciparum* infection.

XX Disclosure; Page 511; 577pp; English.

XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against *P. falciparum* infection comprising (I) or (II); (I) and
CC (II) are useful for the development of vaccines against *P. falciparum*
CC infection. (1) and polyclonal antisera or a monoclonal antibody raised to
CC immunogens comprising the sequences of (I), are useful in the detection
CC of infection with *P. falciparum*. Furthermore, (I) (especially when they
CC are refined or secreted or membrane proteins) can aid the identification
CC of drugs to treat or prevent *P. falciparum* infection, or they can be used
CC to identify drug resistance in *P. falciparum*. Sequencing of the
CC Plasmodium chromosome 2 and the subsequent identification of proteins
CC encoded by it will help to expand our understanding of parasite biology,
CC a process hampered by the complexity of the parasitic lifecycle, and
CC provide new targets for vaccine and drug development. Parasite resistance
CC to drugs and mosquito resistance to insecticides have led to a resurgence
CC of malaria in many parts of the world, and there is a pressing need for
CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
CC represent nucleotide and protein sequences given in the present
CC invention, but which are not specifically mentioned within the
CC specification

XX Sequence 3738 BP; 1705 A; 258 C; 542 G; 1233 T; 0 U; 0 Other;

Query Match 5.3%; Score 45.2; DB 3; Length 3738;
Best Local Similarity 51.5%; Pred. No. 0.28;
Matches 104; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 192 TAATACACGATTTGTAATCCGATGAACTGCAATTAATAAAGTGGCGTCTGATTT 251

DB 2730 TAAGACAATGAAAAATAACCGAATTAACAATAAATAATATATGATATGACATAT 2789

QY 252 AGAAATGGTATTATTCTGCTGATTAAGAGATCCTTATATGATATATGATATGATATGC 311

DB 2790 TGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATG 2849

QY 312 TTCTCACTTTTATGATCCGATATCTGGAACAACATATATCTCTTTTGGAAACATGCAAA 371

DB 2850 TGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATG 2909

QY 372 AGAAACAGCGCAAAATATTTT 393

DB 2910 AGACTTATATGTAACAAATTTT 2931

RESULT 9

ADA71938

ID ADA71938 standard; DNA; 2000 BP.

XX ADA71938;

AC ADA71938;

XX 20-NOV-2003 (first entry)

DE Rice gene, SEQ ID 5263.

XX Plant; bacterial infection; fungal infection; viral infection; rice;
XX gene; ds.

OS Oryza sativa.


```
RESULT 11
ACF74582/c
ID ACF74582 standard; DNA, 837 BP.
XX
XX
AC ACF74582;
XX
XX
DT 20-NOV-2003 (first entry)
XX
XX
DE Staphylococcus aureus DNA #2262.
XX
XX
KW Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;
KW enzymatic assay; antibiotic target; gene; ds.
XX
XX
OS Staphylococcus aureus.
XX
XX
PN WO200294868-A2.
XX
XX
PD 28-NOV-2002.
XX
XX
PF 27-MAR-2002; 2002WO-IB002637.
XX
XX
PR 27-MAR-2001; 2001GB-00007661.
XX
XX
PA (CHIR-) CHIRON SPA.
XX
XX
PI Masignani V, Mora M, Scarselli M;
XX
XX
DR WPI; 2003-120786/11.
DR P-PSDB; ABM73022.
XX
XX
PT New Staphylococcus aureus protein, useful as a vaccine for treating or
PT preventing Staphylococcal infection, specifically an infection caused by
PT S. aureus, e.g. sepsis.
XX
XX
PS Claim 6; SEQ ID NO 4523; 49pp; English.
XX
XX
CC The invention relates to novel genes and encoded proteins from
CC Staphylococcus aureus. A composition comprising the S. aureus protein, a
CC nucleic acid encoding the protein, or an antibody to the protein, is
CC useful as a pharmaceutical, particularly as a vaccine for treating or
CC preventing infection due to Staphylococcus bacteria, specifically an
CC infection caused by S. aureus. The composition is particularly useful for
CC treating or preventing sepsis in a patient. The composition can also be
CC used for diagnostics. The protein is also used in an assay for enzymatic
CC studies and as a target for antibiotics. This sequence represents one of
CC the novel S. aureus genes of the invention
XX
SQ Sequence 837 BP; 314 A; 116 C; 149 G; 258 T; 0 U; 0 Other;
Query Match 5.1%; Score 43.2; DB 7; Length 837;
Best Local Similarity 50.5%; Pred. No. 0.58;
Matches 105; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
QY 403 GGTCAAGCATACAAAATCAAGATATGACGAGCAATCTTCTACTTAGGATTATCGCTT 462
Db 820 GGCCAAATCGACCTTGACCAGTATGTGCAGAAATACCTGGTGTAGTTACATTTATCGT 761
QY 463 CATTTATTAGGAGATGTAATCAGCAATGATGACGCAAACTTTACGAATCTTTCTTAT 522
Db 760 AATTATTACATGAAAGCTTCATTTAAAACTTTCTTCAATTTATCAACATATTCATGA 701
QY 523 CCAATGGGTTTCCATTTCTAAATACGAAATTTTGTGTATACAAATAAAATAACTATTT 582
Db 700 CGTTAGCATGTGGACACCAATGGATTGATTTCAATCTCTATATAATTCAGCAATTT 641
QY 583 GTTTCAGATAGCAATGGATTTGGAATT 610
Db 640 CCTTTTCAAGTATTGGATAGTAATT 613
RESULT 12
ADC21339/c
ID ADC21339 standard; DNA, 837 BP.
XX
XX
AC ADC21339;
XX
XX
DT 18-DEC-2003 (first entry)
XX
XX
DE Staphylococcus aureus protein coding sequence SEQ ID NO: 21.
XX
XX
KW ds; gene; antibacterial; gene therapy; vaccine; diagnosis;
KW microbial survival; gene function; bacterial infection;
KW antibiotic-resistant bacterium.
XX
XX
OS Staphylococcus aureus.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..837
FT /tag= a
XX
XX
PN WO2003029484-A2.
XX
XX
PD 10-APR-2003.
XX
XX
PF 28-SEP-2001; 2001WO-US030379.
XX
XX
PR 28-SEP-2001; 2001WO-US030379.
XX
XX
PA (PHRA ) PHARMACIA & UPJOHN CO.
XX
XX
PI Tomich CC, Quinn CL, Arvidson S, Mott JE, Harris DW;
XX
XX
DR WPI; 2003-372002/35.
DR P-PSDB; ADC21340.
XX
XX
PT Identifying an agent that binds a polypeptide and useful for treating or
PT preventing staphylococcal infections, comprises contacting a polypeptide
PT and an agent to form a mixture and determining whether the agent binds
PT the polypeptide.
XX
XX
Example 1; SEQ ID NO 21; 174pp; English.
XX
XX
CC The invention relates to methods of identifying an agent that binds a
CC polypeptide from Staphylococcus aureus by contacting a polypeptide and an
CC agent to form a mixture and determining whether the agent binds the
CC polypeptide. The method is useful in identifying agents that bind gene
CC products critical for the survival of microbes, preferably staphylococcus
CC microbes, including agents that interfere with the function of such gene
CC products. The method is used to identify new agents useful for treating
CC or preventing bacterial infections, particularly those caused by
CC antibiotic-resistant bacteria. This sequence represents the coding region
CC for a protein from Staphylococcus aureus of the invention.
XX
SQ Sequence 837 BP; 314 A; 116 C; 149 G; 258 T; 0 U; 0 Other;
Query Match 5.1%; Score 43.2; DB 9; Length 837;
Best Local Similarity 50.5%; Pred. No. 0.58;
Matches 105; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
QY 403 GGTCAAGCATACAAAATCAAGATATGACGAGCAATCTTCTACTTAGGATTATCGCTT 462
Db 820 GGCCAAATCGACCTTGACCAGTATGTGCAGAAATACCTGGTGTAGTTACATTTATCGT 761
QY 463 CATTTATTAGGAGATGTAATCAGCAATGATGACGCAAACTTTACGAATCTTTCTTAT 522
Db 760 AATTATTACATGAAAGCTTCATTTAAAACTTTCTTCAATTTATCAACATATTCATGA 701
QY 523 CCAATGGGTTTCCATTTCTAAATACGAAATTTTGTGTATACAAATAAAATAACTATTT 582
Db 700 CGTTAGCATGTGGACACCAATGGATTGATTTCAATCTCTATATAATTCAGCAATTT 641
QY 583 GTTTCAGATAGCAATGGATTTGGAATT 610
Db 640 CCTTTTCAAGTATTGGATAGTAATT 613
```

```

RESULT 13
ADC25025/c
ID   ADC25025 standard; DNA; 837 BP.
XX
XX
AC   ADC25025;
XX
XX
DT   18-DEC-2003 (first entry)
XX
XX
DE   DNA encoding S. aureus polypeptide #11.
XX
XX
KW   Binding agent; growth rate; microbe; virulence; vaccine;
KW   bacterial infection; antibiotic-resistant bacteria; bacteraemia;
KW   septic shock; metastatic infection; endocarditis; arthritis;
KW   osteomyelitis; pneumonia; abscess; skin rash; food poisoning;
KW   multisystem dysfunction; toxic shock syndrome; antimicrobial;
KW   antibacterial; immunosuppressive; cytostatic; antiinflammatory; gene; ds.
XX
XX
OS   Staphylococcus aureus.
XX
XX
FN   US2003087321-A1.
XX
XX
PD   08-MAY-2003.
XX
XX
PF   28-SEP-2001; 2001US-00966521.
XX
XX
PR   28-SEP-2001; 2001US-00966521.
XX
XX
PA   (TOMI/) TOMICH C C.
PA   (QUIN/) QUINN C L.
PA   (ARVI/) ARVIDSON S.
PA   (MOTT/) MOTT J E.
PA   (HARR/) HARRIS D W.
XX
XX
PI   Tomich CC, Quinn CL, Arvidson S, Mott JE, Harris DW;
XX
XX
WPI; 2003-606127/57.
XX
XX
P-ESDB; ADC25026.
XX
XX
PT   Identification of agent that binds polypeptide, for treating bacterial
PT   infections, particularly caused by antibiotic-resistant bacteria,
PT   involves mixing polypeptide and agent, and determining whether agent
PT   binds polypeptide.
XX
XX
PS   Claim 49; SEQ ID NO 21; 122pp; English.
XX
XX
CC   The invention discloses a method for identifying an agent that binds a
CC   polypeptide which comprises mixing a polypeptide and an agent to form a
CC   mixture and determining whether the agent binds the polypeptide. Also
CC   claimed is a method for identifying an agent that decreases the growth
CC   rate of a microbe, making a Staphylococcus aureus with reduced virulence,
CC   which comprises altering a coding sequence in an S. aureus to introduce a
CC   mutation, a vaccine composition comprising the S. aureus organism with
CC   reduced virulence. The methods are useful for identifying an agent that
CC   binds a polypeptide, useful in treating (e.g. as a vaccine) bacterial
CC   infections, particularly those caused by antibiotic-resistant bacteria,
CC   particularly bacteraemia, septic shock and serious metastatic infections
CC   including endocarditis, arthritis, osteomyelitis, pneumonia, abscesses in
CC   virtually any organ, skin rashes, food poisoning or multisystem
CC   dysfunction, i.e. toxic shock syndrome. The inventive method identifies
CC   agents that are unrelated to existing antimicrobials and that target
CC   different aspects of staphylococcal invasion of and replication in the
CC   host. The sequence presented is a DNA sequence encoding one of the S.
CC   aureus polypeptides of the invention.
XX
XX
SQ   Sequence 837 BP; 314 A; 116 C; 149 G; 258 T; 0 U; 0 Other;

Query Match          5.1%; Score 43.2; DB 9; Length 837;
Best Local Similarity 50.5%; Pred. No. 0.58;
Matches 105; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY   403 GGTCAGCATACCAAAATCAAGATATGCGCAAGCAATCTTCTACTTAGGATTATCGCTT 462
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB   820 GGCCAAATGCGACCTTGACCATATGTCGAGAAATAAAGTGGTGTAGTTACATTATATCGT 761
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY   463 CATTATTAGGAGATGTGAATCAGCAATGATGCGCAAACTTTTACGAATCTTCTTAT 522
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 760 AATTATTACATGAAAGGCTTCATTAAAACTTTCTTCAATTTATCAACATATTCATGA 701
 QY 523 CCAATGGGTTTCCATCTTAATACGAAATTTTGGTGTATACATATAAAATACTATATT 582
 Db 700 CTTAGCATGTCGACACCAATGGATTGATTTCAATGATCTCTTATAAATTCAGCAATTT 641
 QY 583 GTTTCAGATAGCAATGGATATTGGAAAT 610
 Db 640 CTTTTCAGATATTGGATAGATT 613

RESULT 15

AAA70220
 ID AAA70220 standard; DNA; 2700 BP.
 XX
 AC AAA70220;
 XX
 XX 07-NOV-2000 (first entry)
 XX
 XX Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:353.
 XX
 DE Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW antimalarial; malaria; protozoa; infection; insecticide; ds.
 XX
 OS Plasmodium falciparum.
 XX
 XX WO200025728-A2.
 XX
 PD 11-MAY-2000.
 XX
 XX 05-NOV-1999; 99WO-US026796.
 PF
 XX 05-NOV-1998; 98US-0107131P.
 PR
 XX (HOFF/) HOFFMAN S.
 PA (CARU/) CARUCCI D.
 PA (GARD/) GARDNER M.
 PA (VENT/) VENTER J C.
 XX

PI Hoffman S, Carucci D, Gardner M, Venter JC;

DR WPI; 2000-365347/31.

XX
 XX Proteins encoded by chromosome 2 of the human malarial parasite,
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 PT diagnosis of P.falciparum infection.

XX
 PS Disclosure; Page 542-543; 577pp; English.

XX The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (I) or (II). (I) and
 CC (II) are useful for the development of vaccines against P. falciparum
 CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to
 CC immunogens comprising the sequences of (I), are useful in the detection
 CC of infection with P. falciparum. Furthermore, (I) (especially when they
 CC are xifins or secreted or membrane proteins) can aid the identification
 CC of drugs to treat or prevent P. falciparum infection, or they can be used
 CC to identify drug resistance in P. falciparum. Sequencing of the
 CC Plasmodium chromosome 2 and the subsequent identification of proteins
 CC encoded by it will help to expand our understanding of parasite biology,
 CC a process hampered by the complexity of the parasitic lifecycle, and
 CC provide new targets for vaccine and drug development. Parasite resistance
 CC to drugs and mosquito resistance to insecticides have led to a resurgence
 CC of malaria in many parts of the world, and there is a pressing need for
 CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
 CC represent nucleotide and protein sequences given in the present
 CC invention, but which are not specifically mentioned within the
 CC specification

XX Sequence 2700 BP; 1152 A; 270 C; 390 G; 888 T; 0 U; 0 Other;

Query Match
 Best Local Similarity 5.1%; Score 43.2; DB 3; Length 2700;
 Matches 200; Conservative 0; Mismatches 0; Indels 6; Gaps 2;
 QY 192 TAATACACAGATTGTGAATCCGAATCGAATGAACTGCAATTAATAATGAGTGGCGTCTGATTT 251
 Db 294 TTATGATAGGATAAATAATTTCTGAAGAAGAAACAAACATTAAACGATGATAATAATGATGA 353
 QY 252 AGAAAAATGGTATTATTCTGCTGATTAAGGAAATCCTTATTATGATAATAGTACATATGC 311
 Db 354 TAATAATGGTGAATTATGATGATGATAATAATAGTATGATGATGATGATGATGATAA 413
 QY 312 TTCTCACATTTTATGATCGGATCTG-----GAACAACATATATTTCTTTTTCGGAACAT 366
 Db 414 TAATAACAATGATGATAATAATAATGATGATGATGATGATGATGATGATGATGATGAT 473
 QY 367 GCAAAAGAAACAGGCGCAAAATATTTTAACTTGTGTCAGGCATACCAAAATCAAGAT 426
 Db 474 AAAAGAAAATGATGAATATAAAGATCTCTATTTCAGATATCTATATAAAGAAAGCAAAAA 533
 QY 427 ATGCAGCAGCATTTCTTCTACTTAGGATTAATCGCTTCATTATTAGGAGATGTGA-ATCA 485
 Db 534 ATGCAATATACGTTGTGAAAAATATTTATGAACAGTTCGGTTTAAATAAGAAGAAATTTAGAAGA 593
 QY 486 GCCAATGATCGCAGCAAACTTTTACGAATCTTTCTTATCCAATGGTTTCCATTCTTAAATA 545
 Db 594 AATAAATGAATCGGATCCCTGACAGTTTCAGATAATAGTATGATCTCATCATCAGAAGA 653
 QY 546 CGAAAAATTTTGTGATACAAATPAAAAAATACTATATTGTTTTCAGATAGCAATGGATATTG 605
 Db 654 AAGTTGTTCTGAGGAATCAGATAAAGAATCAGATAAAGAATCAGATAAAGAATGGAATTTT 713
 QY 606 GAAT 609
 Db 714 ATAT 717

Search completed: August 2, 2004, 02:22:00
 Job time : 436 secs

GenCore version 5.1.6
'Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2004, 02:02:34 ; Search time 87 Seconds
(without alignments)
5415.555 Million cell updates/sec

Title: PCT-US03-12556-1

Perfect score: 849
Sequence: 1 atgaaaagaagtattagc.....atcgtatgtaaatcgtaa 849

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	47.4	5.6	7218	1	US-08-232-463-14
C 2	46.6	5.5	3706	4	US-08-913-159-9
C 3	43.4	5.1	9821	4	US-08-956-171E-470
C 4	43.2	5.1	5924	4	US-08-956-171E-130
C 5	42	4.9	1872	4	US-09-543-681A-455
C 6	41.6	4.9	2911	4	US-08-961-527-199
C 7	39.8	4.7	1113	1	US-08-341-538A-1
C 8	39.8	4.7	1113	2	US-08-725-538-1
C 9	39.8	4.7	1353	4	US-09-601-198-37
C 10	39.2	4.6	832	4	US-09-621-976-2813
C 11	39	4.6	640681	4	US-09-790-988-1
C 12	38.4	4.5	352	3	US-08-961-083-67
C 13	38.4	4.5	352	4	US-09-536-784-67
C 14	38.4	4.5	10207	1	US-08-920-822-2
C 15	38.4	4.5	10207	1	US-08-920-827-2
C 16	38.4	4.5	10207	1	US-08-921-177-2
C 17	38.4	4.5	10207	1	US-08-362-577C-2
C 18	38.4	4.5	10207	2	US-08-920-828-2
C 19	38.4	4.5	15766	3	US-09-338-907-73
C 20	38.4	4.5	15766	4	US-09-218-207-73
C 21	38.4	4.5	37950	3	US-09-338-907-183
C 22	38.4	4.5	37950	4	US-09-218-207-183
C 23	38.2	4.5	2172	4	US-09-976-594-125
C 24	38	4.5	1254	4	US-09-134-001C-973
C 25	38	4.5	2649	4	US-09-565-909-1
C 26	38	4.5	5050	4	US-09-565-909-6
C 27	38	4.5	5128	4	US-09-565-909-5

28	37.8	4.5	832	4	US-09-621-976-2813	Sequence 2813, Ap
29	37.6	4.4	426	4	US-09-328-352-2011	Sequence 2011, Ap
30	37.4	4.4	1269	4	US-09-601-198-178	Sequence 178, App
31	37.2	4.4	579	4	US-09-134-001C-1835	Sequence 1835, Ap
32	37	4.4	765	4	US-08-914-479A-3	Sequence 3, Appli
33	37	4.4	855	4	US-08-914-479A-5	Sequence 5, Appli
34	37	4.4	918	3	US-08-937-271-9	Sequence 9, Appli
35	37	4.4	1332	2	US-08-795-475-5	Sequence 5, Appli
36	37	4.4	1617	4	US-09-328-352-2404	Sequence 2404, Ap
37	36.8	4.3	4594	4	US-08-956-171E-200	Sequence 200, App
38	36.8	4.3	7172	4	US-08-961-527-120	Sequence 120, App
39	36.6	4.3	2820	4	US-09-134-000C-2179	Sequence 2179, Ap
40	36.2	4.3	1664976	4	US-08-916-421B-1	Sequence 1, Appli
41	35.6	4.2	740	3	US-09-451-117-1	Sequence 1, Appli
42	35.6	4.2	740	4	US-09-888-655-1	Sequence 1, Appli
43	35.6	4.2	1950	4	US-09-134-001C-1054	Sequence 1054, Ap
44	35.6	4.2	2042	2	US-08-911-434A-2	Sequence 2, Appli
45	35.6	4.2	3006	4	US-09-762-724-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 5.6%; Score 47.4; DB 1; Length 7218;

```
Best Local Similarity 4.7%; Pred. No. 0.0033;
Matches 12; Conservative 151; Mismatches 92; Indels 0; Gaps 0;

QY 541 AAATACGAAATTTGTTGATACATAAATAAATACTATATGTTTCAGATGCAATGGA 600
Db 1453 AGATAGAAGAAATTTGTTGATACRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTR 1394
QY 601 TATTGGAATTTGGAAGGACCAACCCAGAGATTGGATTGAAGGAGCAGCGTAGCAGT 660
Db 1393 RRRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTR 1334
QY 661 AAACAGATTTATCTGCGGTTTGTAACGATACACAAAGAGATTGGTTTGTAAAGCAGC 720
Db 1333 RRRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTR 1274
QY 721 GTATCTCAAGATATGCAATAAATGCGGTGCGGGAAGTAACACCGGTGACAGGAAGCGT 780
Db 1273 RRRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTR 1214
QY 781 TTAATGGAAGCGCAG 795
Db 1213 RRRRTTTRRTTTRRTTTR 1199

RESULT 2
US-08-913-159-9/c
; Sequence 9, Application US/08913159
; Patent No. 6300109
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Plasmid-derived type II
; TITLE OF INVENTION: restriction-modification systems from Lactococcus lactis
; NUMBER OF SEQUENCES: 14
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,159
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 0179/95
; FILING DATE: 17-FEB-1995
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3706 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Lactococcus lactis subsp. cremoris
; STRAIN: W56
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (422..2161)
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 422
; OTHER INFORMATION: /product= "LlaBI methylase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "orf"
; OTHER INFORMATION: /number= 1
; OTHER INFORMATION: /standard_name= "Gene coding for LlaBI methylase"
; OTHER INFORMATION: /label= m-llaBI
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2464..3360
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 2464
; OTHER INFORMATION: /product= "LlaBI endonuclease"
```

```
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "orf"
; OTHER INFORMATION: /number= 2
; OTHER INFORMATION: /standard_name= "Gene coding for LlaBI
; OTHER INFORMATION: endonuclease"
; OTHER INFORMATION: /label= r-llaBI
US-08-913-159-9

Query Match 5.5%; Score 46.6; DB 4; Length 3706;
Best Local Similarity 47.7%; Pred. No. 0.0041;
Matches 136; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 51 AGTTCAAAGTCTAGTATTGTCACAAACAATAAATAGTGAAGTCTGCGGATTTTAAG 110
Db 1010 ATTTAATCGAATAATATATCTTATGAAATATGAAAGCAATGCTGACCAATCAGCAT 951
QY 111 ATGCTCAGCTGAGGATGAAGCATAATGAGGGGATTAACTCTCATTTGTGGATTGAAATCG 170
Db 950 TGTCTCAATGAAGGAAAAATTCCTTAATCTTTAAATTAATTTATTTGCTATTAAAGACGA 891
QY 171 TGCATTTGACATCATGTCTGTATATACACAGATTGTGAATCCGAACTGCAATATT 230
Db 890 GCTCAGTAAAGCGCAATGCTGCGGAAACCAACCGTTAACATTTGGTATGAAATGCGCATC 831
QY 231 AAATGAGTGGCGTCTGATTTAGAAAAATGGTATTATTCTCTGCTGATTACGAGAACTCCTTA 290
Db 830 TCAAGCAATTGAATAGCCATGTACCTAAATAATTTTCTTACAAATTTCTTTGAAATCCAAA 771
QY 291 TTATGATAATAGTACATATGCTTCTCATTCTTTATGATCCGGATAC 335
Db 770 CTTTGTATATTTTACAGATTATGCTCTTTTAAATAATGATATGC 726

RESULT 3
US-08-956-171E-470
; Sequence 470, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
```

TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 470:
SEQUENCE CHARACTERISTICS:
LENGTH: 9821 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 470:
US-08-956-171E-470

Query Match 5.1%; Score 43.4; DB 4; Length 9821;
Best Local Similarity 44.3%; Pred. No. 0.045;
Matches 176; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

QY 321 TTATGATCCGGTACTGGAACAACATATATCTCTTTTGGCAACATGCAAAAGAAAGACAGG 380
DB 4223 TGATGGTCAGATTTTCAGAAATGAAAGAAAGAAATTTGAAAGCTTGAATCAAGAGC 4282
QY 381 CGCAAAATATTTTAACCTTGTGTCAGCATACCAAAATCAAGATATGCGCAAGCAAT 440
DB 4283 TGACATCACTGTGTAAGAATTTGAGTAAGACTGAAAAGAGCAAGAGCGTATTTTAGTAAG 4342
QY 441 CTCTACTTAGGATATCGCTTCATTATTAGGAGATGTGAATCAGCAATGCGATGCGAGC 500
DB 4343 AATGCAAGAAACAGAAATGCTTATCAATAGCAGCAAGCAAGCAATTTAAAGAGC 4402
QY 501 AAACCTTTACGAATCTTCTTATCCATGGGTTTCCATTTCTAAATACGAAAATTTTGTGA 560
DB 4403 AGAAAAGCAAGAAAGCAAGAAAGAAAGAGTGGTAAGCAGTATGAAGATGATGTCAT 4462
QY 561 TACAATAAAATACTATTTGTTTCAGATAGCAATGGATTTGGAAATTTGGAAGAGGAGC 620
DB 4463 TGCTATAAAATAACGTCACCTTTCTAAAGTCTGAAAAGAGATAAATTTGTAGCTATTGC 4522
QY 621 AAACCCAGAGATTCGATTGAGGAGCAGCGTAGCAGTACAGTACCAAGATTTCCCTGGCGT 680
DB 4523 TGATCAAGACATTAAGGATGAAGTAAGAAAGCAAAATCTAAAAGAGTGTGTAGTAGA 4582
QY 681 TGTGAACGATAGCAGCAAAAGATTTGTTGTAAAGCA 717
DB 4583 CGTTGTTAAAGCAAAATAAGATTTGATTAAGAA 4619

RESULT 4
US-08-956-171E-130/c
; Sequence 130, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248PL
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 130:
SEQUENCE CHARACTERISTICS:
LENGTH: 5924 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 130:
US-08-956-171E-130

Query Match 5.1%; Score 43.2; DB 4; Length 5924;
Best Local Similarity 50.5%; Pred. No. 0.043;
Matches 105; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 403 GGTCAAGCATACCAAAATCAAGATATGCAAGCAAGATTTCTTCTACTTAGGATATCGCTT 462
DB 5152 GGCCTAATCGACCTTGACCAATGTCAGAAATAAAGTGGTGTAGTTACATTTATATCGT 5093
QY 463 CATTTATTAGGAGATGCAATCAGCAATGATGCAAGCAATTTTACGAATCTTCTTAT 522
DB 5092 AATTATTCATGAAAGCTTCATTAATAAATTTCTTCAATTTATCAATATTCATGA 5033
QY 523 CCAATGGGTTTCCATTTCTAAATACGAAAATTTTGTGTATACAAATAAAAAAATAACTATAT 582
DB 5032 CGTTAGCATGTGCGACCAATGGATTTGATTTTCATGATCTCCTATAAATTCAGCAATTT 4973
QY 593 GTTTCAGATAGCAATGGATATTTGGAAT 610
DB 4972 CTTTTCAGTATTTGGATCTACTMGAAT 4945

RESULT 5
US-09-543-681A-455
; Sequence 455, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 455
; LENGTH: 1872
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-455

Query Match 4.9%; Score 42; DB 4; Length 1872;
Best Local Similarity 47.7%; Pred. No. 0.057;
Matches 123; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 350 TTCCTTTTGGCAACATGCAAAAGAAAGCAAGCGGCAAAATATTTTACCTTCTGCTGCTCAAG 409
DB 1612 TTACTGCTGTTTGAGAGAAAAGATAAAGAAAGAAAGATTCACCTTGTGTTTTTAA 1671
QY 410 CATACCAAAATCAAGATATGCAAGCAAGCATTTCTTCTACTTAGGATATTCGCTTCAATAT 469
DB 1672 CAGTCATTAAGCAAGATTTCTTACATCATGCTGCTATGCTAGGTTATTTT 1731

ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 352 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-536-784-67

Query Match 4.5%; Score 38.4; DB 3; Length 352;
Best Local Similarity 44.6%; Pred. No. 0.28;
Matches 150; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

QY 244 GCTGATTAGAAAATGGTATTATTCTGCTGATTACGAGAAATCCTTATTATGATAATAGT 303
DB 8 GGTGAATTAGAAAAGGATACCAATTTGATGGTGGGAAATTTCTGGTTTCGAAGGTAAA 67
QY 304 ACATATGCTTCTCACTTTTATGATCCGGATCTGGAACAAATATATTCCTTTTCGGAAA 363
DB 68 AAGACGCTGGCTATGTTTATTAATCTATCAAAAGATACCTTTTATAAAACCTGTATTCAAG 127
QY 364 CATGCAAAAGAAACAGGCGCAAAATATTTTAACCTTGTGCTCAAGCATACCAAAATCAA 423
DB 128 AAATAGAGGAGAAAAGAGAGAGAAATTAACCTACTTTTGTATCGAAAAGAAA 187
QY 424 GATATGCAAGCAAGCTTCTTCTAGGATTATCGCTTCAATTTATTAGGAGATGTAAT 483
DB 188 GATAACCCACAGTAAACCATAGTCAATTAATGAAGTCACAGAAAAGAGGATTTACAA 247
QY 484 CAGCAATGCTGCGCAAACTTTTACGAATCTTTCTTATCCAAATGGGTTTCCATTTCAA 543
DB 248 AGAGAAGAGCATTCACAAAATCTGATTCAACTAAGGATGTTACAGCTACAGTCTCTGAT 307
QY 544 TACGAAAATTTGTTGATACATAAATAAATCACTAT 579
DB 308 AAAACAATATCAGTAGTAATCACTACTACAAAT 343

RESULT 13
US-09-536-784-67
Sequence 67, Application US/09536784
Patent No. 6573082
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997

ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 352 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-536-784-67

Query Match 4.5%; Score 38.4; DB 4; Length 352;
Best Local Similarity 44.6%; Pred. No. 0.28;
Matches 150; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

QY 244 GCTGATTAGAAAATGGTATTATTCTGCTGATTACGAGAAATCCTTATTATGATAATAGT 303
DB 8 GGTGAATTAGAAAAGGATACCAATTTGATGGTGGGAAATTTCTGGTTTCGAAGGTAAA 67
QY 304 ACATATGCTTCTCACTTTTATGATCCGGATCTGGAACAAATATATTCCTTTTCGGAAA 363
DB 68 AAGACGCTGGCTATGTTTATTAATCTATCAAAAGATACCTTTTATAAAACCTGTATTCAAG 127
QY 364 CATGCAAAAGAAACAGGCGCAAAATATTTTAACCTTGTGCTCAAGCATACCAAAATCAA 423
DB 128 AAATAGAGGAGAAAAGAGAGAGAAATTAACCTACTTTTGTATCGAAAAGAAA 187
QY 424 GATATGCAAGCAAGCTTCTTCTAGGATTATCGCTTCAATTTATTAGGAGATGTAAT 483
DB 188 GATAACCCACAGTAAACCATAGTCAATTAATGAAGTCACAGAAAAGAGGATTTACAA 247
QY 484 CAGCAATGCTGCGCAAACTTTTACGAATCTTTCTTATCCAAATGGGTTTCCATTTCAA 543
DB 248 AGAGAAGAGCATTCACAAAATCTGATTCAACTAAGGATGTTACAGCTACAGTCTCTGAT 307
QY 544 TACGAAAATTTGTTGATACATAAATAAATCACTAT 579
DB 308 AAAACAATATCAGTAGTAATCACTACTACAAAT 343

RESULT 14
US-09-536-784-67
Sequence 2, Application US/08920812
Patent No. 5763188
GENERAL INFORMATION:
APPLICANT: Ohno, Teuneya
APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/920,812
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: US 08/362,577
/ FILING DATE: 27-MAR-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Rin-Laures, Li-Hsien
/ REGISTRATION NUMBER: 33,547
/ REFERENCE/DOCKET NUMBER: 19036/32420
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 312/474-6300
/ TELEFAX: 312/474-0448
/ TELEX: 25-3856
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 10207 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: Genomic DNA
/ ORIGINAL SOURCE:
/ ORGANISM: Staphylococcus aureus
/ STRAIN: Clinical Isolate SA-24
/ US-08-920-812-2

Query Match 4.5%; Score 38.4; DB 1; Length 10207;
Best Local Similarity 50.0%; Pred. No. 1.1;
Matches 96; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 204 TGTGAATCCGAATGAACTGCATTATTAAATGAGTCGCGTCTGATTAGAAAATGGTAT 263
Db 3818 TGCTAAGCAACCAAAATACATTATTGAGCAATTCGCTGCTTGGTTAGATGATGAAT 3877

QY 264 TTATTCGCTGATTACGAGATCCTTATTATGATAATAGTACATATGCTTCTCATTTTA 323
Db 3878 TGTGCTGATAGTTGGTTAGATGCTTATCAAGTAATTTAGATAGCGATCAATTTAAATCA 3937

QY 324 TGATCCGGATACGGAACACATATTCCTTTGCGAAACATGCAAAAGAACAGCGGC 383
Db 3938 AGGTTAGATTATTAAATGTCAGCAATTAAGCTTTGACAACTGTAATAATTAGGTGA 3997

QY 384 AAAATATTTAA 395
Db 3998 AACGTTGCTAA 4009

RESULT 15
US-08-920-827-2
/ Sequence 2, Application US/08920827
/ Patent No. 5770375
/ GENERAL INFORMATION:
/ APPLICANT: Ohno, Tsuneva
/ APPLICANT: Matsuhisa, Akio
/ APPLICANT: Uehara, Hirotugu
/ APPLICANT: Eda, Soji
/ TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
/ NUMBER OF SEQUENCES: 25
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
/ STREET: 6300 Sears Tower, 233 South Wacker Drive
/ CITY: Chicago
/ STATE: Illinois
/ COUNTRY: United States of America
/ ZIP: 60606-6402
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent in Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/920,827
/ FILING DATE: 29-AUG-1997
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/362,577
/ FILING DATE: 27-MAR-1995

/ ATTORNEY/AGENT INFORMATION:
/ NAME: Rin-Laures, Li-Hsien
/ REGISTRATION NUMBER: 33,547
/ REFERENCE/DOCKET NUMBER: 19036/32420
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 312/474-6300
/ TELEFAX: 312/474-0448
/ TELEX: 25-3856
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 10207 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: Genomic DNA
/ ORIGINAL SOURCE:
/ ORGANISM: Staphylococcus aureus
/ STRAIN: Clinical Isolate SA-24
/ US-08-920-827-2

Query Match 4.5%; Score 38.4; DB 1; Length 10207;
Best Local Similarity 50.0%; Pred. No. 1.1;
Matches 96; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 204 TGTGAATCCGAATGAACTGCATTATTAAATGAGTCGCGTCTGATTAGAAAATGGTAT 263
Db 3818 TGCTAAGCAACCAAAATACATTATTGAGCAATTCGCTGCTTGGTTAGATGATGAAT 3877

QY 264 TTATTCGCTGATTACGAGATCCTTATTATGATAATAGTACATATGCTTCTCATTTTA 323
Db 3878 TGTGCTGATAGTTGGTTAGATGCTTATCAAGTAATTTAGATAGCGATCAATTTAAATCA 3937

QY 324 TGATCCGGATACGGAACACATATTCCTTTGCGAAACATGCAAAAGAACAGCGGC 383
Db 3938 AGGTTAGATTATTAAATGTCAGCAATTAAGCTTTGACAACTGTAATAATTAGGTGA 3997

QY 384 AAAATATTTAA 395
Db 3998 AACGTTGCTAA 4009

Search completed: August 2, 2004, 04:12:30
Job time : 102 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2004, 01:48:00 ; Search time 2814 Seconds

(without alignments)
9009.592 Million cell updates/sec

Title: PCT-US03-12556-1

Perfect score: 849

Sequence: 1 atgaaaagaagattagc.....atacgtatgtaaatcgctaa 849

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pln:*
- 20: em_gss_vrt:*
- 21: em_gss_fun:*
- 22: em_gss_mam:*
- 23: em_gss_nus:*
- 24: em_gss_pro:*
- 25: em_gss_rod:*
- 26: em_gss_phg:*
- 27: em_gss_vrl:*
- 28: gb_gss1:*
- 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	52.2	6.1	994	13	BX414650
c 2	51.8	6.1	844	29	BX139987
c 3	51.4	6.1	605	12	BMI62501
c 4	50.4	5.9	1101	29	CNS00396

c 5	50	5.9	977	28	AZ687764	AZ687764	ENTHDB9TF
c 6	49.2	5.8	634	12	BMI64871	BMI64871	EST567394
c 7	48.2	5.7	776	28	BZ063272	BZ063272	1k269h12
c 8	48	5.7	754	12	BX376728	BX376728	BX376728
c 9	47.8	5.6	1201	13	BX355654	BX355654	BX355654
c 10	47	5.6	862	28	AZ677978	AZ677978	ENTHJ38TR
c 11	47	5.5	1101	29	CNS0172Q	AL108704	Drosophil
c 12	46.8	5.5	1190	29	CNS020N7	AL206908	Tetraodon
c 13	46.4	5.5	686	29	CNS03412	AL226880	Tetraodon
c 14	46	5.4	1201	13	BX361152	BX361152	BX361152
c 15	46	5.4	1201	13	BX462896	BX462896	BX462896
c 16	45	5.3	357	12	B1814749	B1814749	PFBST0a40
c 17	45	5.3	1175	28	BZ696793	BZ696793	SP_Ba008
c 18	44.8	5.3	1201	9	AL536104	AL536104	AL536104
c 19	44.6	5.3	404	28	AZ140343	AZ140343	SP_0001 A
c 20	44.6	5.3	740	29	CNS02YR2	AL220040	Tetraodon
c 21	44.6	5.3	843	29	CNS00CS1	AL059666	Drosophil
c 22	44.4	5.2	541	29	CNS04KUS	AL295309	Tetraodon
c 23	44.2	5.2	833	28	BH154104	BH154104	ENTRBS8TF
c 24	44.2	5.2	1101	29	CNS00LT2	AL078714	Drosophil
c 25	44	5.2	613	12	CF181891	CF181891	ISO2H2F I
c 26	44	5.2	666	14	BJ401356	BJ401356	BJ401356
c 27	44	5.2	829	29	CNS00A4K	AL054917	Drosophil
c 28	44	5.2	838	28	AZ195507	AZ195507	SP_1030 A
c 29	44	5.2	1056	13	BX415058	BX415058	BX415058
c 30	44	5.2	1124	13	BX436282	BX436282	BX436282
c 31	44	5.2	1200	13	BX437758	BX437758	BX437758
c 32	43.8	5.2	575	28	BZ780846	BZ780846	1i20h11.g
c 33	43.8	5.2	705	12	BJ446775	BJ446775	BJ446775
c 34	43.8	5.2	1201	13	BX461310	BX461310	BX461310
c 35	43.6	5.1	429	9	AU262155	AU262155	AU262155
c 36	43.6	5.1	494	28	AZ179729	AZ179729	SP_0166 A
c 37	43.6	5.1	526	9	AU037906	AU037906	AU037906
c 38	43.6	5.1	543	29	BX233514	BX233514	Danio rer
c 39	43.6	5.1	816	12	BJ437362	BJ437362	BJ437362
c 40	43.6	5.1	859	29	CNS00KLL	AL077728	Drosophil
c 41	43.6	5.1	992	29	CNS06P6W	AL408990	T7 end of
c 42	43.4	5.1	1201	9	AL536104	AL536104	AL536104
c 43	43.2	5.1	470	9	AU034262	AU034262	AU034262
c 44	43.2	5.1	609	12	BJ444006	BJ444006	BJ444006
c 45	43.2	5.1	1101	29	CNS016LI	AL106896	Drosophil

ALIGNMENTS

RESULT 1

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BX414650/c
BX414650 Homo sapiens THYMUS Homo sapiens cdna clone CS0CAP001YN02
3-PRIME, mRNA sequence.
BX414650
BX414650.1 GI:30763455
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 994)
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6015.f
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0CAP001D001NP1.
Location/Qualifiers
1. .994

FEATURES

source

Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 844)
Humphray, S.J., Huckle, E. and Durham, J.L.
Direct Submission
Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Unpublished
This sequence was generated from the T7 end of BAC 99E7. 99E7 is
part of the Daniokey BAC Library created by R. Plasterk and N.V.
keygene. Further details:
http://www.sanger.ac.uk/Projects/D_rerio/
source
1. .844
Location/Qualifiers
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-99E7"
/tissue_type="Testis"
/note="vector pindigoBAC-536"
ORIGIN
Query Match 6.1%; Score 51.8; DB 29; Length 844;
Best Local Similarity 47.2%; Pred. No. 0.029;
Matches 158; Conservative 0; Mismatches 177; Indels 0; Gaps 0;
Qy 68 TTGCACAAACAAATAGTGAAGTCTCTCCACCGATTTTAAAGATGGTCACTGAGGATA 127
Db 507 TTGATGATAATAATAATGATAATAATAATAATGATAATAATAATGATAATAATA 566
Qy 128 AGCAATATGAGGGATTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 187
Db 567 ATAATAATGATAATAATAATGATAATAATAATGATAATAATAATGATAATAATA 626
Qy 188 CTCTGATATACAGGATGTCGATCCGATCCGATCCGATCCGATCCGATCCGATCCG 247
Db 627 ACAATAATAATAATAATGATAATAATAATAATGATAATAATAATGATAATAATA 686
Qy 248 ATTTAGAAAATGGTATTTATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 307
Db 687 ATGATAATAATAATAATGATAATAATAATAATGATAATAATAATGATAATAATA 746
Qy 308 ATGCTTCTCATTCTTATGATCCGATCTGGAACACATATATCTCTCTCTCTCTCT 367
Db 747 ATAATGATAATAATAATGATAATAATAATGATAATAATAATGATAATAATAATA 806
Qy 368 CAAAGAAAACAGGCGCAAAATATTTTAACTTGTCT 402
Db 807 ATAATAATGATAATAATAATAATAATAATAATGAT 841
RESULT 3
BM162501 605 bp mRNA linear EST 04-DEC-2001
LOCUS EST565024 PyBS Plasmodium yoelii yoelii cDNA clone PYCKT40 5' end,
DEFINITION mRNA sequence.
ACCESSION BM162501
VERSION BM162501.1 GI:17308182
SOURCE EST.
ORGANISM Plasmodium yoelii yoelii
Plasmodium yoelii yoelii
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 605)
Carlton, J.M., Dally, T.M., Long, C.A., Bergman, L.W., Vaidya, A.B.,
Fraser, C.M. and Carucci, D.J.
Plasmodium yoelii EST project at TIGR
Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSOAP001YN02"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
ORIGIN
Query Match 6.1%; Score 52.2; DB 13; Length 994;
Best Local Similarity 20.7%; Pred. NO. 0.024;
Matches 125; Conservative 202; Mismatches 276; Indels 1; Gaps 1;
Qy 195 TACAACGATGTGATCGAATCGAATGCAATTAATAATGAGTGGCGTCTGATTTAGA 254
Db 970 WAAADRWWAGNARRADWDADWAAWWRWRTAWAARKKWADWTTTKTKKXAD 911
Qy 255 AAATGGTATTTATCTGCTGATTACGAGATCTCTTATTATGATATGATGATGATGCTTC 314
Db 910 AADWGDWDWDATAWMAADWMAAARWAAATWMTDWTAWTKDKATWKADDDADDAK 851
Qy 315 TCATTTTATGATCGGATCTGGAAACACATATATCTCTTTTGGAAACATGCAAAAGA 374
Db 850 TKTITTTTTTWWKAWTDWAGTAKWADATWADADWAAARWAAARWRAAAWMAAD 791
Qy 375 AACAGCGCAAAATATTTTAACTTGTCTGTCGATCAAGCATACCAAAATCAAGATATG 434
Db 790 ADWNNWWDADAAWAAWDTKRWRAADWABAAAAAANWWDWGRGRADAGKDKAR 731
Qy 435 AGCAT-TCTTCTACTAGGATATCGCTTCTATTTATTAGGAGATGTGATCGCAATGC 493
Db 730 AAAMDAGGGRDKDRDWAARWDDADAAARAAAAAATAWATWTKWKKDKMAAAAWDTA 671
Qy 494 ATGAGCAAACTTACGATCTCTTCTTATCCATGGTTTCCATCTTAAATACGAAATTT 553
Db 670 KGKGRKDDARDWMAWTTTTTTTTTAAWAAATTKTKTKTKTKTKTKTKTKTKTKTK 611
Qy 554 TTGTTGATACAAATAAATACTATATTTTTCAGATAGCAATGATATGGAATGGA 613
Db 610 KTTAKWTDWADWTTTTTTTTTTTTTAAWAAATTKTKTKTKTKTKTKTKTKTKTKTK 551
Qy 614 AAGGACAAACCGAAGATTGGATTGAAGGAGCAGCGTACGATCAACAGATATTC 673
Db 550 AAFARRAARRAN 491
Qy 674 CTGGCTGTGACGATACGACAAAGATTGGTTTGTAAAGCAGCGCTATCTCAAGAT 733
Db 490 NRRDRNRKNNTKRWATWTKRGRTKGTGTTTTTTTTTTTDDTKDTGGGKGTRAGAAA 431
Qy 734 ATGCAGATAAATGGCGTGGAGTAAACACCGGTGACAGAAAGCGTTTAAATGGAAGCG 793
Db 430 AGRRKDAAMGRRGKGGAGGAWAAARAAAAARAAAGAGAAAGAAATTAATAAAATAGG 371
Qy 794 AGCG 797
Db 370 RGCG 367
RESULT 2
BX139987 844 bp DNA linear GSS 13-MAR-2003
LOCUS Danio rerio genomic clone DKEY-99E7, genomic survey sequence.
DEFINITION BX139987
ACCESSION BX139987
VERSION BX139987.1 GI:127971314
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Db 800 RKBADKDDAARDRAATTTTTRTDDDKMKTDTWRAADRTDRDDDDDR 741

Qy 503 ACTTTACGAATCTTTCTTATCAATGGTGGTTCCATCTTAAATAGCAAAATTTTGTGATA 562

Db 740 AGTAGKWRRTWKRRKRDTRWDADDDARDRRRRGGDADAGKGGKTKGRKRRRD 681

Qy 563 CAATAAAAATAACTATATTTGTTTCAGATAGCAATGGATATTGGAATTTGGAAGAGGACAA 622

Db 680 RATWDTDAWADAANWTTTDDTDDDKRBRKRGARRRTTARAANDWTTWKANDWA 621

Qy 623 ACCCAGAAGATTGGATTGAAGAGAGCGGCTAGCAGCTTAAACAAGATTATCTCGCGCTG 682

Db 620 KWDKTRADRWDRMAADTTWDARKADRWAKARAWARRDRARAARDRTWTTGKKTITA 561

Qy 683 TGAACGATACGACAAAAGATTGGTTT 708

Db 560 TTTTWAABAANWAWAWATTATTTT 535

RESULT 5

AZ687764/c

LOCUS

DEFINITION

ENTHD89TF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 977)

Lofus, B., Van Aken, S. and Fraser, C.

Determination of clone end sequences from Entamoeba histolytica HM1:IMSS sheared DNA library

Unpublished (2000)

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: b.loftus@tigr.org

Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library

Seq primer: M13-Forward

Class: shotgun

High quality sequence start: 58

High quality sequence stop: 493.

Location/Qualifiers

1. .977

/organism="Entamoeba histolytica"

/mol_type="genomic DNA"

/strain="HM1:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/notes="Vector: pHO31; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

FEATURES

source

Query Match 5.9%; Score 50; DB 28; Length 977;

Best Local Similarity 49.4%; Pred. No. 0.084;

Matches 157; Conservative 0; Mismatches 160; Indels 1; Gaps 1;

Qy 375 AACAGGCGCAAAATATTTTAACTCTGCTCAAGCATACCAAAATCAAGATATGACGA 434

Db 560 ATCAAGTGCAAAAGATTTGTCATGATGAAGGGTTAAACAAATAAGTTTACACCA 501

Qy 435 AGCAATCTTCTACTTAGGATTATCGCTTCATTA-TTTAGGAGATGTGAATCAGCAATGC 493

Db 500 TTGAAGAAGTTATTAATAATATGCAAAATTCATTAGTGGAAACGGTTGCACCAACCTCCTT 441

Qy 494 ATCAGCAAACTTTAGCAATCTTTTATCCAAATGGTTTCCCATTTCTAAATACGAAAT 553

Db 440 ATCCAAAATAGCCTTCATCTCTTCCTATCCATATCATATTCCTGTATGTACCACTG 381

Qy 554 TTGTTGATACATAAAAAATACTATATTTGTTTTCAGATAGCAATGATATTTGGAATGGA 613

Db 380 GTGTTGATTATTAGAAAAATGCTGGATGGAATAGAGATATTACTGGGTACAGTTCAG 321

Qy 614 AAGAGCAAAACCCAGAAGATTGGATTGAAGAGCAGCGGTAGCAGCTTAACAAGATTATC 673

Db 320 ATGAAGTAAAGAAGACAAACCAAGAAAGAACCTGAGTTTGAAGTTGATAATGATAATG 261

Qy 674 CTGCGGTTGTGAACGATA 691

Db 260 ATGTTGTTCTTATGGA 243

RESULT 6

BM164871

LOCUS

DEFINITION

EST567394 PyBS Plasmodium yoelii yoelii cDNA clone PYCMD07 5' end, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 634)

Fraser, C.M. and Carucci, D.J.

Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Vaidya, A.B., Plasmodium yoelii EST project at TIGR

Unpublished (2001)

Contact: Jane Carlton

Parasite Genomics Group

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-530-9319

Fax: 301-538-0208

Email: carlton@tigr.org

For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC

http://www.malaria.mr4.org/mr4pages/index.html

Seq primer: ADF.

Location/Qualifiers

1. .634

/organism="Plasmodium yoelii yoelii"

/mol_type="mRNA"

/strain="17XL"

/sub_species="yoelii"

/db_xref="taxon:73239"

/clone="PYCMD07"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven

FEATURES

source

termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybrizAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the HybrizAP vector and plasmid DNA isolated."

ORIGIN

Query Match 5.8%; Score 49.2; DB 12; Length 634;
Best Local Similarity 50.9%; Pred. No. 0.12;
Matches 117; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 547 GAAATTTTGTGATACATAAATAAATACTATATTTCTTCAGATAGCAATGGATTTGG 606
DB 95 GAAAGTTTGGAGGAAAAATAAAAAAGATAAAAATGGGTTATGAGAAAAATGAAAGTCAA 154
QY 607 AATTGGAAGGAGCAACCCAGCAAGATTGGATTGAAGGAGCGGTAGCAGCTAAACAA 666
DB 155 GATAGTACAGAGAAAGTCAAGATGAAGAGATGAAGAGGATGAAGAGAGAA 214
QY 667 GATTATCTGGCGTTGTGAACATACGACAAAGAGATTGTTTGTAAAAAGCAGCCGTATCT 726
DB 215 GATGAAGATGAAGAGATGAAGAGAGAAAGATGATCTCTGAAAAATTAAGATTAATT 274
QY 727 CAAGAAATATGAGATAAATGGCGTGCAGGAAGTAACACCGGTGACAGGAAA 776
DB 275 AAAGAAGAACTATAGATGTCCACCATTAGCTCAATATAGAGAGAGA 324

RESULT 7

BZ063272/c
LOCUS lk269hl2.b1 B.oleracea002 Brassica oleracea genomic, genomic survey sequence.
DEFINITION BZ063272.1 GI:23673911

ACCESSION

VERSION BZ063272.1

KEYWORDS

SOURCE GSS.

ORGANISM

Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE

1 (bases 1 to 776)
Delehaunty,K., Fewell,G., Fulton,L., McComb,W.R., Miner,T.,

AUTHORS

Nash,W., Rabinowicz,P.D. and Wilson,R.K.

TITLE

Whole genome shotgun reads from Brassica oleracea

JOURNAL

COMMENT

Unpublished (2002)
Contact: Richard K. Wilson

Genome Sequencing Center
Washington University School of Medicine

Email: submissions@watson.wustl.edu

Plate: lk269 row: h column: 12

Seq primer: -21UpPOT forward

Class: shotgun

High quality sequence start: 8

High quality sequence stop: 551.

FEATURES

source

1..776
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"

/note="Vector: pOTW13; Whole genome shotgun library from flowering buds. DNA was purified from a crude nuclear prep using Brassica oleracea T0100DH3 buds provided by Thomas Osborn at the University of Wisconsin. Genomic DNA was provided by Pablo Rabinowicz (CSHL) and the shotgun library prepared at Washington University Genome Sequencing Center."

ORIGIN

Query Match 5.7%; Score 48.2; DB 28; Length 776;
Best Local Similarity 50.7%; Pred. No. 0.23;
Matches 113; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 268 TCTGCTGATTACGAGAATCCCTTATTATGATAATAGTACATATGCTTCTCACCCTTTATGAT 327
DB 750 TCTTATTATGTCGAGAACAGTATTTCTTATTATCAAGTTGTATCTTATTATTATTAN 691
QY 328 CCGGATACTGGAAACAACATATATTCCTTTTTCGAAAAACATGCAAAAAGAACAGGCCAAAA 387
DB 690 AATTATACCTGCTTCTGGAGTACATAAAGCTTCTTAAAGAGATAAATAAATGAAGTAAAT 631
QY 388 TATTTTAACCTTGTGCTCAAGCATACCAAAATCAAGATATGCGACAGGATCTTCTTAC 447
DB 630 AATATTAGTCTGGAAGAAAAATAGAAAAAATAAATAAATTTGAATATACAAATNTAGTG 571
QY 448 TTAGGATTATCGCTTTCATTATTTAGGAGATGTGAATCAGCCAA 490
DB 570 GGGTATTAGGACTTAATTTTGAAGAGGCCCATTTATTGNA 528

RESULT 8
BZ063272/c
LOCUS BZ063272.1 GI:19286111

DEFINITION BZ063272
Dictyostelium discoideum cDNA clone ddc29o13 3', mRNA sequence.

ACCESSION BZ063272
VERSION BZ063272.1
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.

REFERENCE 1 (bases 1 to 754)
Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
Full length cDNA of Dictyostelium discoideum at the culmination stage
Unpublished (2002)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
Location/Qualifiers
source
1..754
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc29o13"
/sex="mat A"
/dev_stage="Culmination stage"
/clone_lib="Dictyostelium discoideum cDNA library, CF"

ORIGIN

Query Match 5.7%; Score 48; DB 12; Length 754;
Best Local Similarity 45.7%; Pred. No. 0.26;
Matches 168; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

QY 193 AATCAACAGATTGGAATCCGAATGAAACTGCATTATTAATGAGTGGCGTGCTGATT 252
DB 487 AATAAAATGGGTGATGATTATGATGAAGAAATTTAATAAGAAATTTAATAAGAAATTA 428
QY 253 GAAATGGTATTATTCGCTGATTACGAGATCTTATTATGATTAATAGTACATGCT 312
DB 427 GATATTACAGTAGTAGTGAAGAATATGATGATTAATTAATAATAAATCAATGATGAT 368
QY 313 TCTCATTATGATCCGATACCTGGAACAACATATATTCCTTTTGGAAAAATGCAAAA 372
DB 367 GATGATGATATGATGAAGATGATGCTGGAATATTATCAAGAAGAGGATTGAAGAT 308
QY 373 GAAACAGCGCCAAAATATTTTAACTTGGTGGTCAAGCATACCAAAATCAAGATATGAG 432

```

Db      307 GAAAAACCTGAAATTCAAATTAATAAATAATTTGAAGGATGATCATGATGATGAT 248
QY      433 CAAGCATCTCTACTTACTAGGATATCCGCTCATTTATTAGAGATGGAATCAGCCAAAG 492
Db      247 GATGATTTATTATGATGAAGGAGGAGAAATTTGTTGAATATTATGATAAAATGTAATATA 188
QY      493 CATGACAGCAAACTTTAGCAATCTTCTTCTATCCATGGGTTTCCATCTTAATACGAAAT 552
Db      187 GTTGATATAATGATCCAAATGTTGAATATATTCAAGGTGATGATGATGATGATGATAAT 128
QY      553 TTGTTGGA 560
Db      127 GATAATGA 120

```

```

RESULT 9
LOCUS   BX355654
DEFINITION BX355654 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CSODI002YE12 3-PRIME, mRNA sequence.
ACCESSION BX355654
VERSION   BX355654.1 GI:30371982
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens

```

```

REFERENCE 1 (bases 1 to 1201)
AUTHORS   Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Genoscope - Centre National de Sequencage

```

```

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODI002BC06NP1.
Location/Qualifiers
1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODI002YE12"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

```

```

FEATURES
source

```

```

ORIGIN

```

```

Query Match      5.6%; Score 47.8; DB 13; Length 1201;
Best Local Similarity 28.6%; Pred. No. 0.31;
Matches 131; Conservative 110; Mismatches 214; Indels 3; Gaps 1;

QY      385 AAATATTTAACTTGGTCTGCTCAAGCATACCAAAATCAAGATATGCAGCAGCATTTCTC 444
Db      523 ATATATTTATTTTATTAATTTGTTAATAAATAAATTAATTAATAAATAAATAAATTA 582
QY      445 TACTTAGGATATCGCTTCATTTTGGAGATGATGATGATGATGATGATGATGATGATGAT 504
Db      583 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 642
QY      505 TTTAGCAATCTTCTTCTTCAATGGGTTTCCATCTCAAAATACGAAATTTTGTGTATCA 564
Db      643 DAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 702
QY      565 ATAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 624
Db      703 WAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 759

```

```

QY      625 CCAGAAGATTGGATTGAAGGAGCAGCGGTAGCAGTAAACAAGATTATCTCTGGCGTGTG 684
Db      760 AAKKAAATAAAGAAKADKGAADBDWKGAATAAATAAATAAATAAATAAATAAATAAATAA 819
QY      685 AACGATAGACAAAGATGTTGTTGTAATAAAGACCGGTATCTCAAGATATGCGAGATAAA 744
Db      820 GKADGAAATAAGGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 879
QY      745 TGGCGTGGCGGAAGTAACACCGGTGACAGGAAGCGTTTAAATGGAAGCGCGCGGTATCA 804
Db      880 KGKGGKKKKKAAADKDAADAGAAAKKAAATAAATAAATAAATAAATAAATAAATAAATAA 939
QY      805 GCTGTTATATTCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 842
Db      940 ARADKGAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 977

```

```

RESULT 10
LOCUS   AZ677978
DEFINITION ENTHJ38TR Entamoeba histolytica Sheared DNA linear GSS 14-DEC-2000
genomic, genomic survey sequence.
ACCESSION AZ677978
VERSION   AZ677978.1 GI:11815244
KEYWORDS Entamoeba histolytica
SOURCE   Entamoeba histolytica
ORGANISM Eukaryota; Entamoebidae; Entamoeba.

```

```

REFERENCE 1 (bases 1 to 862)
AUTHORS   Loftus, B., Van Aken, S. and Fraser, C.
TITLE     Determination of clone end sequences from Entamoeba histolytica
JOURNAL   HMI:IMSS sheared DNA library
COMMENT   Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjl@loftusetgr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 15
High quality sequence stop: 862.

```

```

FEATURES
source

```

```

1. .862
Location/Qualifiers
/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHS01; Site 1: Bat I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
```

```

ORIGIN

```

```

Query Match      5.5%; Score 47; DB 28; Length 862;
Best Local Similarity 48.9%; Pred. No. 0.46;
Matches 154; Conservative 0; Mismatches 160; Indels 1; Gaps 1;

```

QY 378 AGCGCAAAATATTTAACTTGGTCAAGCATACCAAAATCAAGATATCGCAAGC 437
DB 26 AAGTGCAAAAGATTGTGATGAATGAAGGTTTAAACAAATTAACGTTTATACATTG 85
QY 438 ATTCTTCTACTTACTGAGTATCGCTTCATTA-TTTAGGAGATGTGAATCAGCAATGCATG 496
DB 86 AAGAGTTTATTAATATGCAAAATTCATTTAGTGGACGGTTCACACAACTCTTATC 145
QY 497 CAGCAAACTTTTACGAATCTTCTTATCCAAATGGTTCCTTCAATCAAGAAATTTTG 556
DB 146 CAAAAATGCTTCATCTCTTCTTATCCATATCATATTCCTGTTAATGTACCACTGGTG 205
QY 557 TTGATACATAAATAAATACTATATTTTTCAGATAGCAATGGATTTGGATTTGGAAG 616
DB 206 TTGATTTTATTAAGAAATGCTGGAATAGAGATATTAATCTGCTACAGTGAATCAGATG 265
QY 617 GAGCAAAACCCAGAAATGGATTTGAAGGAGCAGCGGTAGCAGCTAAACAAAGATATCTCGT 676
DB 266 AAAGTAAAGAGAACACCAAGAAAGAACCTGAGTTTGAAGTTGATTAATGATATGATG 325
QY 677 GCGTTGTGAACGATA 691
DB 326 TTGTTCTTATGAAA 340

RESULT 11
CNS0172Q 1101 bp DNA linear GSS 26-JUL-1999
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN37P08 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION
VERSION AL108704.1 GI:5629008
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (23-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk-. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES
source
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN37P08"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/note="end : SP6"

ORIGIN
Query Match 5.5%; Score 47; DB 29; Length 1101;
Best Local Similarity 29.6%; Pred. No. 0.48;
Matches 112; Conservative 81; Mismatches 185; Indels 0; Gaps 0;

QY 225 ATTATTAATAGTGGCGTGTGATTAGAAATGGTATTTATCTGCTGATTACGAGA 284
DB 724 AHKYNKWTGTGHHGGYGGKXTHHGGKGGTAANGKAMGTGCHDARWCCGKTCM 783
QY 285 TCCTTTATGATAATAGTACATATGCTTCTCACTTTTATGATCCGGATCTGGAACAAC 344

DB 784 KASNAAMCAYACKYRAAAAWAAHWMTKHTYCTTTWATATMYMTGRTMTTAARAAMKRM 843
QY 345 ATATATTTCTTTTGGCAAAACATGCAAAAGAACAGCGCAAAATATTTTAACTTCTGCTGG 404
DB 844 DNKAKATVATATAWAAATATATTTTAAWATATWMTKTYTYYTBTBTATWWTAT 903
QY 405 TCAAGCATACCAAAATCAAGATATGCAAGCATCTTCTTCTAGGATATACGCTTCA 464
DB 904 ATATTTATWATATWATATATTTTMMWMTWAAWAAATTTMMWMTATWATATATATATAW 963
QY 465 TTATTTAGGAGATGTGAATCAGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 524
DB 964 ATATTTAAWATATATAWATAAATAAATAATATWMTATATATATATATATWMTAAWMTAT 1023
QY 525 AATGGTTTCCATCTTAAATACGAAATTTTGTTCATACAAATAAAAAAATCAATATATGTT 584
DB 1024 WATTATTTAT 1083
QY 585 TTCAGATAGCAATGGATA 602
DB 1084 WTTWMTWATWMTAWATA 1101

RESULT 12
CNS020N7/c 1190 bp DNA linear GSS 01-SEP-2000
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
153P04 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION
VERSION AL206908.1 GI:7865727
KEYWORDS
SOURCE
ORGANISM
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontoidea; Tetraodontidae; Tetraodon.

REFERENCE
1
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)

2
20296633
10835845
Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)

3 (bases 1 to 1190)
Genoscope.
Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES
source
1..1190
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="153P04"

```

/clone lib="G"
/notes="Genoscope sequence ID : COAG153DH02SP1-end :
PUC-Ori"

ORIGIN
Query Match      5.5%; Score 46.8; DB 29; Length 1190;
Best Local Similarity 36.6%; Pred. No. 0.54;
Matches 132; Conservative 37; Mismatches 192; Indels 0; Gaps 0;

Qy 255 AAATGGTATTTTCTGCTGATTACGAGATCCCTATTATGATATAGTACATATGCTTC 314
Dy 918 ATATAWTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 859
Qy 315 TCACCTTTATGATCGGATCTCGAACCAATATATCTCTTTCGAAACATGCAAAAGA 374
Dy 858 TAAATATATATATAWAAWATATAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWTA 799
Qy 375 AACAGGCGCAAAATATTTTAACTTGTGCTGCTGCAAGCATACCAAAATCAAGATGCGAGCA 434
Dy 798 WTAAWTATATTAATATATATATATATATATATATATATATATATATATATATATATA 739
Qy 435 AGCATTCTTCTACITAGATATCGCTTCATTTATTTAGGAGATGGAATCAGCAATGCA 494
Dy 738 TATTTTATATATATATATATATATATATATATATATATATATATATATATATATATTTT 679
Qy 495 TGCAGCAAACTTTACGAATCTTCTTATCCCAATGGGTTTCCATTCCTAAATACGAAATTT 554
Dy 678 WTTATATATATATATATATATATATATATATATATATATATATATATATATATATTTT 619
Qy 555 TGTGATACAAATAAAAAATACTATTTGTTTCAGATAGCAATGATGATTTGGAATGGAA 614
Dy 618 TWTTTTMMWATWAAATANATATAMAAANNTAWAAWAAWAAWAAWAAWAAWAAWTTTAA 559
Qy 615 A 615
Dy 558 A 558

RESULT 13
CNS0341Z
LOCUS      CNS0341Z      686 bp      DNA      linear      GSS 01-SEP-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            21K19 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION  AL226880
VERSION     AL226880.1  GI:7885817
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
            Tetraodon nigroviridis
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodontidae; Tetraodon.
REFERENCE   1
AUTHORS     Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,P.,
            Saurin,W. and Weissenbach,J.
            Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
JOURNAL     Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE     20296633
PUBMED      10835645
REFERENCE   2
AUTHORS     Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,P.,
            Saurin,W., Bernot,A. and Weissenbach,J.
            Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
JOURNAL     Genome Res. 10 (7), 939-949 (2000)
MEDLINE     20359837
PUBMED      10899143
REFERENCE   3  (bases 1 to 686)
AUTHORS     Genoscope.

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
```

```

Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source            1..686
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="211K19"
                     /clone_lib="G"
                     /notes="Genoscope sequence ID : COAG211AF10SP1-end :
                     PUC-Ori"

ORIGIN
Query Match      5.5%; Score 46.4; DB 29; Length 686;
Best Local Similarity 41.0%; Pred. No. 0.64;
Matches 105; Conservative 29; Mismatches 122; Indels 0; Gaps 0;

Qy 99 ACCGATTTTAAAGATGGTCAGCTGAGGATAGCATATGAGGGGATTAACCTCATTTGTG 158
Dy 200 ACAGGTGAGAAATTTGTTTAACTCAGATAATAATAATAATAATAATAATAATAATA 259
Qy 159 GATTGTAATCGTGCAATTCGACATGTCGCTGTAATACCAACGATTCGTAATCGAATGA 218
Dy 260 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 319
Qy 219 AACTGCATTTATTAATAGTGGGTGCTGATTTAGAAAAATGGTATTTATTCGCTGATTA 278
Dy 320 TAATAATAATAATWATWRTKKTGGTGGTGTDKTKTKTKTKTKTKTKTKTKTKTKTK 379
Qy 279 CGAGATCCCTATTATGATATAGTACATATGCTCTCCTCTTTTATGATCGGATACGG 338
Dy 380 TWATWATGTTTATTAATAATWATWATWATWATWATWATWATWATWATWATWATWAT 439
Qy 339 AACAAACATATATTCCT 354
Dy 440 GTGTACATAGTGTCAT 455

RESULT 14
BX361152/c
LOCUS      BX361152      1201 bp      mRNA      linear      EST 05-MAY-2003
DEFINITION BX361152 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CSODI081YCl9 5-PRIME, mRNA sequence.
ACCESSION  BX361152
VERSION     BX361152.1  GI:30380524
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
AUTHORS     Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
            Full-length cDNA libraries and normalization
            Unpublished (2001)
JOURNAL
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 7593.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CSODI081AB10QPlcluster=7593.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
            Paraday Avenue Genoscope sequence ID : CSODI081AB10QPl.
            Location/Qualifiers

FEATURES
```

```
source
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODM001YL14"
/tissue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoRV sites of the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN
Query Match      5.4%; Score 46; DB 13; Length 1201;
Best Local Similarity 39.1%; Pred. No. 0.87;
Matches 148; Conservative 17; Mismatches 214; Indels 0; Gaps 0;

Qy 226 TTATTAAATGAGTGGCGTCTGATTTAGAAAATGGTATTATTCTGCTGATTACAGAAAT 285
Db 954 TTTTWWAATTKKGGKKTWTWTTTTTAAAAAAAATTTTTTTTTTWWAAAAAAAT 895

Qy 286 CCTTATTATGATAAGTACATATCTCTCACTTTTATGATCCGGATCTGGAACAACA 345
Db 894 TTTTTTTTTTTRARKTTTTTAAAWTTTTTTTTTTTTTTTTTTTTTAAATAATWTTGAA 835

Qy 346 TATATTCCTTTTGGCAACATGCAAAAGAAACAGCGCAAAATATTTTAACTTCTGCTGT 405
Db 834 AAAAAAATTTTTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTT 775

Qy 406 CAAGCATACCAAAATCAAGATATGACGACAGATCTTCTACTAGGATTTATCGCTTCAT 465
Db 774 AWWWWTWWAAWATAAAWAAATAAAWAAAAAATTTTTTTTATTTTTTTTTTTT 715

Qy 466 TATTAGGAGATGTGAATCAGCAATGATGACGCAAACTTTTACGAATCTTCTTATCCA 525
Db 714 TTTTAAARRANNNTNNNNAANAAWAAAAAATTTTTTTTTTTTTTTTTTTTTT 655

Qy 526 ATGGGTTTCCATTCTTAAATACGAAATTTTGTGTGATACAATAAAAA 571
Db 654 TTTTAAATGTAAGTAAAGTTTATTAAATTAATTAAGATCTTGAAAA 609

RESULT 15
BX462896
LOCUS
DEFINITION
BX462896 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
CSODM001YL14 3-PRIME, mRNA sequence.
ACCESSION
BX462896
VERSION
BX462896.1 GI:31019600
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 4458.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODM001DF07NP1&cluster=4458.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODM001DF07NP1.
Location/Qualifiers
1..1201
/organism="Homo sapiens"

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODM001YL14"
/tissue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoRV sites of the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN
Query Match      5.4%; Score 46; DB 13; Length 1201;
Best Local Similarity 39.1%; Pred. No. 0.87;
Matches 148; Conservative 17; Mismatches 214; Indels 0; Gaps 0;

Qy 225 ATTATTAATGAGTGGCGTCTGATTTAGAAAATGGTATTATTCTGCTGATTACAGAA 284
Db 699 AATAWNAATAATTTATTANTATATTTTATAWAATANWATTTTATTANANANAAAAATAA 758

Qy 285 TCCTTATTATGATATAGTACATATGCTTCTCACTTTTATGATCCGGATCTGGAACAAC 344
Db 759 ATAAATTTAAATAAATAATATATATAATTTAAATAAATAAATAAATAAATAAATAA 818

Qy 345 ATATATTCCTTTTGGCAACATGCAAAAGAAACAGCGCAAAATATTTTAACTTCTGCTGT 404
Db 819 TATAATTTATATATAATWAAAAAATTTTAAATATATATAAATAAATAAATAAATAA 878

Qy 405 TCAAGCATACCAAAATCAAGATATGACGACAGATCTTCTACTAGGATTTATCGCTTCA 464
Db 879 TAATTATAATATTTAAATTTAAATAATTTTAAATATATATATAATAATAATAATTNNA 938

Qy 465 TTATTTAGGAGATGTGAATCAGCAATGCAATGCAAGCAAACTTTACGAATCTTCTTATCC 524
Db 939 TTATTTAAATTAATAAANNTAATAATAATAATAATAATAATAATAATAATAATAATTA 998

Qy 525 AATGGTTTCCATTCTTAAATACGAAATTTTGTGTGATACAATAAAAAATACTATATGT 584
Db 999 ATTAATAWTAATAAATWNAATAATAAATAATTTATNTATTTATNTANANATATTTATATA 1058

Qy 585 TTCAGATAGCAATGGATAT 603
Db 1059 WATATANANTATAAATAT 1077

Search completed: August 2, 2004, 04:10:48
Job time : 2820 secs
```

This Page Blank (uspto)